

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**PENETRATION RATE OPTIMIZATION WITH SUPPORT VECTOR
REGRESSION METHOD**

M.Sc. THESIS

Korhan KOR

Department of Petroleum and Natural Gas Engineering

Petroleum and Natural Gas Engineering Programme

JUNE 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**PENETRATION RATE OPTIMIZATION WITH SUPPORT VECTOR
REGRESSION METHOD**

M.Sc. THESIS

**Korhan KOR
(505121501)**

Department of Petroleum and Natural Gas Engineering

Petroleum and Natural Gas Engineering Programme

Thesis Advisor: Asst. Prof. Dr. Gürşat ALTUN

JUNE 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**DESTEK VEKTÖR REGRESYONU YÖNTEMİ İLE İLERLEME HIZI
OPTİMİZASYONU**

YÜKSEK LİSANS TEZİ

**Korhan Kor
(505121501)**

Petrol ve Doğal Gaz Mühendisliği Anabilim Dalı

Petrol ve Doğal Gaz Mühendisliği Programı

Tez Danışmanı: Yrd. Doç. Dr. Gürşat ALTUN

HAZİRAN 2015

Korhan Kor, a M.Sc. student of ITU **Graduate School of Science Engineering and Technology** student ID **505121501**, successfully defended the **thesis** entitled “**PENETRAION RATE OPTIMIZATION WITH SUPPORT VECTOR REGRESSION METHOD**”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Asst. Prof. Dr. Gürsat ALTUN**
Istanbul Technical University

Jury Members : **Asst. Prof. Dr. Şenol YAMANLAR**
Istanbul Technical University

Prof. Dr. Hanifi ÇOPUR
Istanbul Technical University

Date of Submission : 04 May 2015
Date of Defense : 01 June 2015

To my family,

FOREWORD

First of all, I would like to thank to my thesis advisor, Asst. Prof. Dr. Gürşat ALTUN for his endless support during my graduate years as a student, a research assistant and an engineer. I also appreciate him for bringing this unique subject of machine learning to me as my thesis topic.

And the second biggest appreciation is for my mother and father, who have raised me, supported me for education, and made me ready for the tough life starting from my childhood. One of the most important contributors of this thesis is my parents, no doubt. Thank you, for everything.

The university and the department have the most important role on me working as a petroleum and natural gas engineer today. So, I would like to thank individually all the professors in the Department of Petroleum and Natural Gas Engineering in Istanbul Technical University. Also, many thanks to my co-workers in the department during my research assistant years, Ali Ettehadî OSGOUEI, Dr. E. Didem KORKMAZ BAŞEL, Eda AY DİLSİZ, F. Bahar HOŞGÖR, Kağan KUTUN, Dr. M. Hakan ÖZYURTKAN, Melek DENİZ PAKER, and Dr. Yıldırım PALABIYIK.

I also would like to thank to Asst. Prof. Dr. Yusuf YASLAN for letting me listen to his “Machine Learning” lectures and helping me for my questions.

Cengiz BERKÜN, Cenk ŞANLIOĞLU and Mert Can SELÇUK... Thank you for being a big brother for me. Thank you for always being there for me. Do not let your music end. We will always be together.

And finally, Ezgi BİLGİÇ... Thank you for always being on my side, for your huge support, for your big heart, and for your love.

May 2015

Korhan KOR
Petroleum and Natural Gas Engineer

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS.....	xi
ABBREVIATIONS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xix
SYMBOLS	xxiii
SUMMARY	xxv
ÖZET.....	xxvii
1. INTRODUCTION.....	1
2. LITERATURE REVIEW.....	7
2.1 Former Studies on ROP Optimization	7
2.2 Recent Studies on ROP Optimization	17
3. STATEMENT OF THE PROBLEM AND SCOPE OF THE THESIS.....	31
4. MULTIPLE REGRESSION AND BOURGOYNE & YOUNG METHOD... 35	35
4.1 Theory of Multiple Regression Analysis	35
4.1.1 Introduction to multiple regression analysis	35
4.1.2 Least squares estimation of the parameters.....	36
4.1.3 Matrix approach	38
4.2 Bourgoyne and Young Method	39
4.2.1 Effect of formation strength	40
4.2.2 Effect of normal compaction.....	41
4.2.3 Effect of under compaction.....	41
4.2.4 Effect of overbalance	42
4.2.5 Effect of bit weight and bit diameter	43
4.2.6 Effect of rotary speed	44
4.2.7 Effect of tooth wear.....	45
4.2.8 Effect of bit hydraulics.....	47
5. MACHINE LEARNING AND SUPPORT VECTOR REGRESSION.....	51
5.1 Statistical Learning Theory	58
5.1.1 VC dimension	59
5.1.2 Structural risk minimisation.....	60
5.2 Support Vector Machines for Classification	60
5.2.1 The optimal separating hyperplane	62
5.2.2 The generalized optimal separating hyperplane.....	67
5.3 Kernel Functions	69
5.3.1 Linear kernels.....	70
5.3.2 Gaussian radial basis function kernels	71
5.4 Support Vector Regression.....	71
5.4.1 ϵ -insensitive regression	72
5.4.2 ν regression	74

5.5 Cross Validation and Overfitting.....	76
5.5.1 K-fold cross validation	78
5.5.2 Leave-one-out cross validation	78
5.6 Toolboxes	78
5.6.1 R programming language	78
5.6.2 LIBSVM.....	79
5.6.3 e1071 package	79
5.7 Statistical Comparison Criteria.....	79
5.7.1 Residual sum of squares	80
5.7.2 R^2	80
5.7.3 Pseudo- R^2	81
5.7.4 Root mean squared error	81
6. RESULTS AND DISCUSSIONS	83
6.1 Case 1: Training Top 16 Depth Interval.....	84
6.1.1 Testing #17-18-19-20.....	84
6.1.2 Testing #22-23-24-25.....	87
6.1.3 Testing #27-28-29-30.....	90
6.1.4 Testing #17-21-26-30.....	93
6.2 Case 2: Training Bottom 16 Depth Interval	98
6.2.1 Testing #11-12-13-14.....	98
6.2.2 Testing #6-7-8-9.....	101
6.2.3 Testing #1-2-3-4.....	104
6.2.4 Testing #1-5-10-14.....	107
6.3 Case 3: Training Odd Numbered Data Points	113
6.3.1 Testing #2-4-6-8.....	113
6.3.2 Testing #12-14-16-18.....	116
6.3.3 Testing #22-24-26-28.....	119
6.3.4 Testing #2-10-20-30.....	122
6.4 Case 4: Training Even Numbered Data Points.....	127
6.4.1 Testing #1-3-5-7.....	127
6.4.2 Testing #11-13-15-17.....	130
6.4.3 Testing #21-23-25-27.....	133
6.4.4 Testing #1-9-19-29.....	136
6.5 Case 5: Training Several 24 Data Points.....	142
6.5.1 Training first 24, testing last 6 data inputs	142
6.5.2 Training last 24, testing first 6 data inputs	145
6.5.3 Training mid 24, testing #1-2-3-28-29-30.....	148
6.5.4 Training several 24, testing #4-7-10-18-22-27	151
6.6 Case 6: Training All Data Points.....	156
6.6.1 Testing #1-2-3-4-5.....	157
6.6.2 Testing #26-27-28-29-30.....	160
6.6.3 Testing #14-15-16-17-18.....	163
6.6.4 Testing #1-6-11-16-21.....	166
6.6.5 Testing all	171
6.7 Case 7: Training Different Data Set	175
6.8 Discussions and General Interpretations	179
7. CONCLUSIONS AND RECOMMENDATIONS	181
REFERENCES	183
APPENDICES	197
CURRICULUM VITAE.....	209

ABBREVIATIONS

ANN	: Artificial Neural Networks
BYM	: Bourgoyne and Young Model
DENFIS	: Dynamic Evolving Neural-Fuzzy Inference System
GA	: Genetic Algorithm
GDL	: Geological Drilling Log
KKT	: Karush-Kuhn-Tucker
LWD	: Logging While Drilling
ML	: Machine Learning
MSE	: Mechanical Specific Energy
MR	: Multiple Regression
MWD	: Measurement While Drilling
OCSH	: Optimal Canonical Separating Hyperplane
PDC	: Polycrystalline Diamond Compact
PVT	: Pressure Volume Temperature
RBF	: Radial Basis Function
RKHS	: Reproducing Kernel Hilbert Spaces
RMSE	: Root Mean Squared Error
ROP	: Rate of Penetration
RSS	: Residual Sum of Squares
SRM	: Structural Risk Minimization
SV	: Support Vectors
SVM	: Support Vector Machines
SVR	: Support Vector Regression
TCC	: Tungsten Carbide Core
TSS	: Total Sum of Squares
TVD	: Total Vertical Depth
VC	: Vapnik-Chervonenkis
WOB	: Weight on Bit

LIST OF TABLES

	<u>Page</u>
Table 1.1 : Methods of operations research	2
Table 4.1 : Data tabulation for multiple regression.....	36
Table 5.1 : Summary of ML settings	55
Table 6.1 : Multiple regression coefficients of Case 1.....	84
Table 6.2 : Statistics of the data set in Case 1 for testing #17-18-19-20.....	85
Table 6.3 : Statistical results of MR in Case 1 for learning and testing #17-18-19-20.....	85
Table 6.4 : Statistical results of the methods in Case 1 for testing #17-18-19-20. ...	85
Table 6.5 : ROP predictions in Case 1 for #17-18-19-20.....	85
Table 6.6 : Statistics of the data set in Case 1 for testing #22-23-24-25.....	88
Table 6.7 : Statistical results of MR in Case 1 for learning and testing #22-23-24-25.....	88
Table 6.8 : Statistical results of the methods in Case 1 for testing #22-23-24-25	88
Table 6.9 : ROP predictions in Case 1 for #22-23-24-25.....	88
Table 6.10 : Statistics of the data set in Case 1 for testing #27-28-29-30.....	91
Table 6.11 : Statistical results of MR in Case 1 for learning and testing #27-28-29-30	91
Table 6.12 : Statistical results of the methods in Case 1 for testing #27-28-29-30 ..	91
Table 6.13 : ROP predictions in Case 1 for #27-28-29-30	91
Table 6.14 : Statistics of the data set in Case 1 for testing #17-21-26-30.....	94
Table 6.15 : Statistical results of MR in Case 1 for learning and testing #17-21-26-30	94
Table 6.16 : Statistical results of the methods in Case 1 for testing #17-21-26-30 ..	94
Table 6.17 : ROP predictions in Case 1 for #17-21-26-30	94
Table 6.18 : Multiple regression coefficients of Case 2.....	98
Table 6.19 : Statistics of the data set in Case 2 for testing #11-12-13-14.....	99
Table 6.20 : Statistical results of MR in Case 2 for learning and testing #11-12-13-14	99
Table 6.21 : Statistical results of the methods in Case 2 for testing #11-12-13-14	100
Table 6.22 : ROP predictions in Case 2 for #11-12-13-14	100
Table 6.23 : Statistics of the data set in Case 2 for testing #6-7-8-9.....	102
Table 6.24 : Statistical results of MR in Case 2 for learning and testing #6-7-8-9.	102
Table 6.25 : Statistical results of the methods in Case 2 for testing #6-7-8-9	103
Table 6.26 : ROP predictions in Case 2 for #6-7-8-9.....	103
Table 6.27 : Statistics of the data set in Case 2 for testing #1-2-3-4.....	105
Table 6.28 : Statistical results of MR in Case 2 for learning and testing #1-2-3-4.	105
Table 6.29 : Statistical results of the methods in Case 2 for testing #1-2-3-4	106
Table 6.30 : ROP predictions in Case 2 for #1-2-3-4.....	107
Table 6.31 : Statistics of the data set in Case 2 for testing #1-5-10-14.....	108

Table 6.32 : Statistical results of MR in Case 2 for learning and testing #1-5-10-14.....	108
Table 6.33 : Statistical results of the methods in Case 2 for testing #1-5-10-14.....	109
Table 6.34 : ROP predictions in Case 2 for #1-5-10-14.....	109
Table 6.35 : Multiple regression coefficients of Case 3.....	113
Table 6.36 : Statistics of the data set in Case 3 for testing #2-4-6-8.....	114
Table 6.37 : Statistical results of MR in Case 3 for learning and testing #2-4-6-8.....	114
Table 6.38 : Statistical results of the methods in Case 3 for testing #2-4-6-8.....	114
Table 6.39 : ROP predictions in Case 3 for #2-4-6-8.....	114
Table 6.40 : Statistics of the data set in Case 3 for testing #12-14-16-18.....	117
Table 6.41 : Statistical results of MR in Case 3 for testing #12-14-16-18.....	117
Table 6.42 : Statistical results of the methods in Case 3 for testing #12-14-16-18.....	117
Table 6.43 : ROP predictions in Case 3 for #12-14-16-18.....	117
Table 6.44 : Statistics of the data set in Case 3 for testing #22-24-26-28.....	120
Table 6.45 : Statistical results of MR in Case 3 for learning and testing #22-24-26-28.....	120
Table 6.46 : Statistical results of the methods in Case 3 for testing #22-24-26-28.....	120
Table 6.47 : ROP predictions in Case 3 for #22-24-26-28.....	120
Table 6.48 : Statistics of the data set in Case 3 for testing #2-10-20-30.....	123
Table 6.49 : Statistical results of MR in Case 3 for learning and testing #2-10-20-30.....	123
Table 6.50 : Statistical results of the methods in Case 3 for testing #2-10-20-30.....	123
Table 6.51 : ROP predictions in Case 3 for #2-10-20-30.....	123
Table 6.52 : Multiple regression coefficients of Case 4.....	127
Table 6.53 : Statistics of the data set in Case 4 for testing #1-3-5-7.....	128
Table 6.54 : Statistical results of MR in Case 4 for learning and testing #1-3-5-7.....	128
Table 6.55 : Statistical results of the methods in Case 4 for testing #1-3-5-7.....	129
Table 6.56 : ROP predictions in Case 4 for #1-3-5-7.....	129
Table 6.57 : Statistics of the data set in Case 4 for testing #11-13-15-17.....	131
Table 6.58 : Statistical results of MR in Case 4 for testing #11-13-15-17.....	131
Table 6.59 : Statistical results of the methods in Case 4 for testing #11-13-15-17.....	132
Table 6.60 : ROP predictions in Case 4 for #11-13-15-17.....	132
Table 6.61 : Statistics of the data set in Case 4 for testing #21-23-25-27.....	134
Table 6.62 : Statistical results of MR in Case 4 for learning and testing #21-23-25-27.....	134
Table 6.63 : Statistical results of the methods in Case 4 for testing #21-23-25-27.....	135
Table 6.64 : ROP predictions in Case 4 for #21-23-25-27.....	135
Table 6.65 : Statistics of the data set in Case 4 for testing #1-9-19-29.....	137
Table 6.66 : Statistical results of MR in Case 4 for learning and testing #1-9-19-29.....	137
Table 6.67 : Statistical results of the methods in Case 4 for testing #1-9-19-29.....	138
Table 6.68 : ROP predictions in Case 4 for #1-9-19-29.....	138
Table 6.69 : Multiple regression coefficients in Case 5 for testing last 6.....	141
Table 6.70 : Statistics of the data set in Case 5 for testing last 6.....	143
Table 6.71 : Statistical results of MR in Case 5 for learning and testing last 6.....	143
Table 6.72 : Statistical results of the methods in Case 5 for testing last 6.....	143
Table 6.73 : ROP predictions in Case 5 for last 6.....	143
Table 6.74 : Multiple regression coefficients in Case 5 for testing first 6.....	145
Table 6.75 : Statistics of the data set in Case 5 for testing first 6.....	146
Table 6.76 : Statistical results of MR in Case 5 for testing first 6.....	146

Table 6.77 : Statistical results of the methods in Case 5 for testing first 6	146
Table 6.78 : ROP predictions in Case 5 for first 6	146
Table 6.79 : Multiple regression coefficients in Case 5 for testing #1-2-3-28-29-30	148
Table 6.80 : Statistics of the data set in Case 5 for testing #1-2-3-28-29-30	149
Table 6.81 : Statistical results of MR in Case 5 for learning and testing #1-2-3-28-29-30	149
Table 6.82 : Statistical results of the methods in Case 5 for testing #1-2-3-28-29-30	149
Table 6.83 : ROP predictions in Case 5 for #1-2-3-28-29-30	149
Table 6.84 : Multiple regression coefficients in Case 5 for testing #4-7-10-18-22-27	151
Table 6.85 : Statistics of the data set in Case 5 for testing #4-7-10-18-22-27	152
Table 6.86 : Statistical results of MR in Case 5 for learning and testing #4-7-10-18-22-27	152
Table 6.87 : Statistical results of the methods in Case 5 for testing #4-7-10-18-22-27	152
Table 6.88 : ROP predictions in Case 5 for #4-7-10-18-22-27	152
Table 6.89 : Multiple regression coefficients of Case 6	156
Table 6.90 : Statistics of the data set in Case 6 for testing #1-2-3-4-5	157
Table 6.91 : Statistical results of MR in Case 6 for learning and testing #1-2-3-4-5	158
Table 6.92 : Statistical results of the methods in Case 6 for testing #1-2-3-4-5	158
Table 6.93 : ROP predictions in Case 6 for #1-2-3-4-5	158
Table 6.94 : Statistics of the data set in Case 6 for testing #26-27-28-29-30	161
Table 6.95 : Statistical results of MR in Case 6 for learning and testing #26-27-28-29-30	161
Table 6.96 : Statistical results of the methods in Case 6 for testing #26-27-28-29-30	161
Table 6.97 : ROP predictions in Case 6 for #26-27-28-29-30	161
Table 6.98 : Statistics of the data set in Case 6 for testing #14-15-16-17-18	164
Table 6.99 : Statistical results of MR in Case 6 for learning and testing #14-15-16-17-18	164
Table 6.100 : Statistical results of the methods in Case 6 for testing #14-15-16-17-18	164
Table 6.101 : ROP predictions in Case 6 for #14-15-16-17-18	164
Table 6.102 : Statistics of the data set in Case 6 for testing #1-6-11-16-21	167
Table 6.103 : Statistical results of MR in Case 6 for learning and testing #1-6-11-16-21	167
Table 6.104 : Statistical results of the methods in Case 6 for testing #1-6-11-16-21	167
Table 6.105 : ROP predictions in Case 6 for #1-6-11-16-21	167
Table 6.106 : Statistics of the data set in Case 6 for testing all	171
Table 6.107 : Statistical results of MR in Case 6 for learning and testing all	171
Table 6.108 : Statistical results of the methods in Case 6 for testing all	172
Table 6.109 : ROP predictions in Case 6 for all	173
Table 6.110 : Statistics of the data set in Case 7 for testing all	175
Table 6.111 : Multiple regression coefficients in Case 7 for testing all	176
Table 6.112 : Statistical results of MR in Case 7 for learning and testing all	176
Table 6.113 : Statistical results of the methods in Case 7 for testing all	177

Table 6.114 : ROP predictions in Case 7 for all.....	177
Table A.1: The main data set.....	199
Table A.2: The recent data set.....	201

LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : Minimum and maximum of a function.....	1
Figure 1.2 : An example of time and cost relationship	4
Figure 2.1 : Optimum WOB and rotary speed	15
Figure 2.2 : The InterACT system	18
Figure 2.3 : Algorithm of the mathematical optimization procedure	24
Figure 2.4 : Hole cleaning chart example	25
Figure 2.5 : Flowchart of a soft computing approach	26
Figure 2.6 : Decrease in drill bit diameter via wear characteristics	29
Figure 2.7 : Number of papers including SVM found in OnePetro database.	30
Figure 4.1 : Functional relationship of penetration rate.....	40
Figure 4.2 : Effect of normal compaction on ROP	41
Figure 4.3 : Effect of differential bottom-hole pressure on ROP.....	42
Figure 4.4 : Typical response of penetration rate to increasing bit weight	43
Figure 4.5 : Typical response of penetration rate to increasing rotary speed	45
Figure 4.6 : Effect of tooth wear on penetration rate	46
Figure 4.7 : Tooth wear chart for a roller-cone bit.....	47
Figure 4.8 : ROP as a function of bit Reynolds number	47
Figure 4.9 : The effect of different hydraulic models on ROP	48
Figure 5.1 : Expert Systems flowchart.....	52
Figure 5.2 : Genetic Algorithms flowchart	53
Figure 5.3 : Fuzzy Logic flowchart.....	53
Figure 5.4 : Machine Learning related fields	54
Figure 5.5 : An overview of how ML works	54
Figure 5.6 : An example of linear classification in two dimensions.....	55
Figure 5.7 : An example of decision tree	56
Figure 5.8 : Mapping of ML methods.....	57
Figure 5.9 : ML taxonomy	58
Figure 5.10 : Modelling errors	59
Figure 5.11 : VC dimension illustration.....	59
Figure 5.12 : Classifier examples.....	61
Figure 5.13 : Linear classification.....	61
Figure 5.14 : Optimal canonical seperating hyperplane	63
Figure 5.15 : Canonical hyperplanes and constraints	64
Figure 5.16 : Generalized optimal separating hyperplane	67
Figure 5.17 : Feature space illustration	69
Figure 5.18 : Loss functions.....	71
Figure 5.19 : ϵ -insensitive loss function and slack variables	73
Figure 5.20 : Overfitting example.....	76
Figure 5.21 : Reducing overfitting	77
Figure 6.1 : Cost graphics of Case 1 for testing #17-18-19-20.....	84

Figure 6.2 : ROP prediction trends in Case 1 for testing #17-18-19-20	86
Figure 6.3 : ROP prediction comparison in Case 1 for testing #27-28-29-30	86
Figure 6.4 : Cost graphics of Case 1 for testing #22-23-24-25	87
Figure 6.5 : ROP prediction trends in Case 1 for testing #22-23-24-25	89
Figure 6.6 : ROP prediction comparison in Case 1 for testing #22-23-24-25	89
Figure 6.7 : Cost graphics of Case 1 for testing #27-28-29-30	90
Figure 6.8 : ROP prediction trends in Case 1 for testing #27-28-29-30	92
Figure 6.9 : ROP prediction comparison in Case 1 for testing #27-28-29-30	92
Figure 6.10 : Cost graphics of Case 1 for testing #17-21-26-30	93
Figure 6.11 : ROP prediction trends in Case 1 for testing #17-21-26-30	95
Figure 6.12 : ROP prediction comparison in Case 1 for testing #17-21-26-30	95
Figure 6.13 : RSS_{model} values for Case 1	96
Figure 6.14 : CV error values for Case 1	97
Figure 6.15 : Testing errors for Case 1	97
Figure 6.16 : Pseudo- R^2 values for Case 1	98
Figure 6.17 : Cost graphics of Case 2 for testing #11-12-13-14	99
Figure 6.18 : ROP prediction trends in Case 2 for testing #11-12-13-14	99
Figure 6.19 : ROP prediction comparison in Case 2 for testing #11-12-13-14	101
Figure 6.20 : Cost graphics of Case 2 for testing #6-7-8-9	102
Figure 6.21 : ROP prediction trends in Case 2 for testing #6-7-8-9	103
Figure 6.22 : ROP prediction comparison in Case 2 for testing #6-7-8-9	104
Figure 6.23 : Cost graphics of Case 2 for testing #1-2-3-4	105
Figure 6.24 : ROP prediction trends in Case 2 for testing #1-2-3-4	106
Figure 6.25 : ROP prediction comparison in Case 2 for testing #1-2-3-4	107
Figure 6.26 : Cost graphics of Case 2 for testing #1-5-10-14	108
Figure 6.27 : ROP prediction trends in Case 2 for testing #1-5-10-14	109
Figure 6.28 : ROP prediction comparison in Case 2 for testing #1-5-10-14	110
Figure 6.29 : RSS_{model} values for Case 2	111
Figure 6.30 : CV error values for Case 2	111
Figure 6.31 : Testing errors for Case 2	112
Figure 6.32 : Pseudo- R^2 values for Case 2	112
Figure 6.33 : Cost graphics of Case 3 for testing #2-4-6-8	113
Figure 6.34 : ROP prediction trends in Case 3 for testing #2-4-6-8	115
Figure 6.35 : ROP prediction comparison in Case 3 for testing #2-4-6-8	115
Figure 6.36 : Cost graphics of Case 3 for testing #12-14-16-18	116
Figure 6.37 : ROP prediction trends in Case 3 for testing #12-14-16-18	118
Figure 6.38 : ROP prediction comparison in Case 3 for testing #12-14-16-18	118
Figure 6.39 : Cost graphics of Case 3 for testing #22-24-26-28	119
Figure 6.40 : ROP prediction trends in Case 3 for testing #22-24-26-28	121
Figure 6.41 : ROP prediction comparison in Case 3 for testing #22-24-26-28	121
Figure 6.42 : Cost graphics of Case 3 for testing #2-10-20-30	122
Figure 6.43 : ROP prediction trends in Case 3 for testing #2-10-20-30	124
Figure 6.44 : ROP prediction comparison in Case 3 for testing #2-10-20-30	124
Figure 6.45 : RSS_{model} values for Case 3	125
Figure 6.46 : CV error values for Case 3	126
Figure 6.47 : Testing errors for Case 3	126
Figure 6.48 : Pseudo- R^2 values for Case 3	127
Figure 6.49 : Cost graphics of Case 4 for testing #1-3-5-7	128
Figure 6.50 : ROP prediction trends in Case 4 for testing #1-3-5-7	129
Figure 6.51 : ROP prediction comparison in Case 4 for testing #1-3-5-7	130

Figure 6.52 : Cost graphics of Case 4 for testing #11-13-15-17	131
Figure 6.53 : ROP prediction trends in Case 4 for testing #11-13-15-17	132
Figure 6.54 : ROP prediction comparison in Case 4 for testing #11-13-15-17	133
Figure 6.55 : Cost graphics of Case 4 for testing #21-23-25-27	134
Figure 6.56 : ROP prediction trends in Case 4 for testing #21-23-25-27	135
Figure 6.57 : ROP prediction comparison in Case 4 for testing #21-23-25-27	136
Figure 6.58 : Cost graphics of Case 4 for testing #1-9-19-29	137
Figure 6.59 : ROP prediction trends in Case 4 for testing #1-9-19-29	138
Figure 6.60 : ROP prediction comparison in Case 4 for testing #1-9-19-29	139
Figure 6.61 : RSS_{model} values for Case 4	140
Figure 6.62 : CV error values for Case 4	140
Figure 6.63 : Testing errors for Case 4	141
Figure 6.64 : Pseudo- R^2 values for Case 4	141
Figure 6.65 : Cost graphics of Case 5 for testing last 6	142
Figure 6.66 : ROP prediction trends in Case 5 for testing last 6	144
Figure 6.67 : ROP prediction comparison in Case 5 for testing last 6	144
Figure 6.68 : Cost graphics of Case 5 for testing first 6	145
Figure 6.69 : ROP prediction trends in Case 5 for testing first 6	147
Figure 6.70 : ROP prediction comparison in Case 5 for testing first 6	147
Figure 6.71 : Cost graphics of Case 5 for testing #1-2-3-28-29-30	148
Figure 6.72 : ROP prediction trends in Case 5 for testing #1-2-3-28-29-30	150
Figure 6.73 : ROP prediction comparison in Case 5 for testing #1-2-3-28-29-30	150
Figure 6.74 : Cost graphics of Case 5 for testing #4-7-10-18-22-27	151
Figure 6.75 : ROP prediction trends in Case 5 for testing #4-7-10-18-22-27	153
Figure 6.76 : ROP prediction comparison in Case 5 for testing #4-7-10-18-22-27	153
Figure 6.77 : RSS_{model} values for Case 5	154
Figure 6.78 : CV error values for Case 5	155
Figure 6.79 : Testing errors for Case 5	155
Figure 6.80 : Pseudo- R^2 values for Case 5	156
Figure 6.81 : Cost graphics of Case 6 for testing #1-2-3-4-5	157
Figure 6.82 : ROP prediction trends in Case 6 for testing #1-2-3-4-5	159
Figure 6.83 : ROP prediction comparison in Case 6 for testing #1-2-3-4-5	159
Figure 6.84 : Cost graphics of Case 6 for testing #26-27-28-29-30	160
Figure 6.85 : ROP prediction trends in Case 6 for testing #26-27-28-29-30	162
Figure 6.86 : ROP prediction comparison in Case 6 for testing #26-27-28-29-30	162
Figure 6.87 : Cost graphics of Case 6 for testing #14-15-16-17-18	163
Figure 6.88 : ROP prediction trends in Case 6 for testing #14-15-16-17-18	165
Figure 6.89 : ROP prediction comparison in Case 6 for testing #14-15-16-17-18	165
Figure 6.90 : Cost graphics of Case 6 for testing #1-6-11-16-21	166
Figure 6.91 : ROP prediction trends in Case 6 for testing #1-6-11-16-21	168
Figure 6.92 : ROP prediction comparison in Case 6 for testing #1-6-11-16-21	168
Figure 6.93 : RSS_{model} values for Case 6	169
Figure 6.94 : CV error values for Case 6	170
Figure 6.95 : Testing errors for Case 6	170
Figure 6.96 : Pseudo- R^2 values for Case 6	171
Figure 6.97 : Cost graphics of Case 6 for testing all	172
Figure 6.98 : ROP prediction trends in Case 6 for testing all	174
Figure 6.99 : ROP prediction comparison in Case 6 for testing all	175
Figure 6.100 : Cost graphics of Case 7 for testing all	176
Figure 6.101 : ROP prediction trends in Case 7 for testing all	178

Figure 6.102 : ROP prediction comparison in Case 7 for testing all 179

SYMBOLS

A_B	: Borehole area
A_{bed}	: Area of cuttings bed
A_{well}	: Area of the well
A_v	: Projected area of each diamond
A_{vw}	: Projected area of worn section of a diamond
C	: Complexity parameter
C_c	: Cuttings concentration for a stationary bed
C_f	: Formation drillability parameter
D	: Well depth
E	: Rock hardness
E_s	: Specific energy
F_j	: Jet impact force
F_{jm}	: Modified jet impact force
H	: Normalized bit tooth height
K	: Drillability of formation
$K(x, x')$: Kernel function
$L(y, f(x))$: Loss function
M	: Margin
N	: Rotary speed
N_s	: Number of diamond stress
Q	: Volumetric flow rate
R	: Rate of penetration
$R_{emp[f]}$: Empirical risk function
R_n	: Normalized rate of penetration
R^2	: Confidence level
S	: Confined rock strength
S_c	: Compressive strength
T	: Bit torque
V	: Volume of the rock removal
W	: Weight on bit
W_b	: Bit weight
W_f	: Wear function
b	: Bias
d_b	: Bit diameter
d_{be}	: Bearing diameter
d_n	: Bit nozzle diameter
e	: Residual
e_m	: Mechanical efficiency
$f_c(P_e)$: Chip holddown function
g_p	: Pore pressure gradient
h	: Bit tooth dullness
h_{vc}	: VC dimension

l	: Number of training data
m	: Number of insert penetrations per revolution
n_i	: Number of insert in contact with the rock at the bottom
p	: Differential pressure
p_c	: Circulating bottomhole pressure
p_p	: Pore pressure
v_{actual}	: Mud velocity in annulus
$v_{critical}$: Mud critical velocity in annulus
t	: Time
w	: Weight vector
\hat{y}	: Predicted value
\bar{y}	: Mean value
Φ	: Lagrangian
α, β, η	: Lagrange multipliers
γ	: RBF kernel parameter
γ_f	: Fluid specific gravity
ε	: SVR tolerance control parameter
ζ	: Bit-specific coefficient of sliding friction
λ	: Rotary speed exponent
μ	: Plastic viscosity
ρ	: Mud density
ρ_c	: Equivalent circulating density
σ_p	: Ultimate strength of rock at a differential pressure
v	: SVR controller parameter
ν_k	: Kinematic viscosity
ξ	: Slack variables
ψ	: Chip formation angle

PENETRATION RATE OPTIMIZATION WITH SUPPORT VECTOR REGRESSION METHOD

SUMMARY

Drilling operations constitute the major part of the exploration costs. During operations, drill bits are the primary part needs to be changed frequently due to its quick wearing nature. In order to reduce the drilling cost, the optimum bit pulling time must be determined. To determine the optimum bit pulling time, either rate of penetration or the tooth wearing parameter must be estimated. The most common method that developed for estimating the optimum time for bit change is “Bourgoyne and Young” method. In this method, eight parameter coefficients are needed. To obtain these coefficients, thirty different data which can be taken from either different shale zones inside thirty different wells in a field or thirty different shale points from one well is needed. However, when there is not enough data taken from thirty different shale sections, the accuracy of “Bourgoyne and Young” method decreases.

To construct the functional relationship with the data and parameter coefficients, a regression analysis must be performed. In this study, two kind of regression technique is used and the results are compared to each other. First technique is the multiple regression analysis, which is also used in “Bourgoyne and Young” method. This analysis applies least-squares-principled-regression to the data and calculates the parameter coefficients in order to estimate the target function. The second technique is one of different types of machine learning algorithms, called Support Vector Regression. In this technique, first, the data is divided into train and test datasets. Then, the regression model is constructed by using train datasets. At last, the model is applied to test datasets in order to predict the target values for the function.

For the calculations, the selection of training and testing data sets are divided into cases with different scenarios. The results of different predictor methods for each scenario are compared with each other in the corresponding case. The results show the significant effect of data selection on the accuracy of penetration rate prediction.

One of the most powerful methods in machine learning, Support Vector Regression, is used for rate of penetration optimization for the first time in the literature with this thesis study. In this way, the chance for further investigations and studies on the practicability of Support Vector Regression on penetration rate optimization is created.

DESTEK VEKTÖR REGRESYONU YÖNTEMİ İLE İLERLEME HIZI OPTİMİZASYONU

ÖZET

Günümüzde enerji kaynaklarına olan talep artışı nedeniyle petrol, gaz ve jeotermal kaynak arayışları önemini daha da arttırarak korumaktadır. Bu talep artışını karşılamak amacıyla daha önce araştırma yapılmamış yeraltı derinliklerinde ve su derinliğinin 3000 metreyi bulduğu açık denizlerde yeni kaynak araştırmaları devam etmektedir. Bu araştırma giderlerinin büyük bir çoğunluğunu sondaj operasyonları oluşturmaktadır. Doğası gereği, sondaj esnasında kulinin ve sondaj dizisinin en çok aşınmaya uğrayan, çabuk eskiiyip değıştirilmesine ihtiyaç duyulan elemanı matkaptır. Sondajın metraj veya performans maliyetini düşürmek için bir matkabın hem uzun süre çalışması, hem de iyi iş yapması istenir. Matkap çalışıkça aşınacağı için ilerleme hızı azalır ve sondaj maliyeti artmaya başlar. Bu sebepten dolayı, matkabın aşınma durumunun dikkatle takip edilmesi gerekmektedir. Eğer matkap zamanından önce kuyudan çıkarılırsa, başka kuyularda tekrar kullanılma özelliğini çoğunlukla kaybeder. Eğer matkap fazla aşınır ve bu durum fark edilmezse, bazı kısımları (diş, kon, vs.) parçalanarak kuyunun içinde kalır. Bu kalan parçalar çıkarılıp kuyu temizlenmeden sondaja devam edilemeyeceği için tahlisiye olarak adlandırılan kurtarma operasyonlarının yapılmasını zorunlu kılar. Bu durum, zaman ve para kaybına yol açtığı gibi tahlisiye operasyonunun yüzde yüz başarı ile gerçekleşeceğinin garantisi de yoktur.

Sondaj esnasındaki ilerleme hızı birçok parametreye bağlıdır. Bu sebeple, ilerleme hızını tahmin veya optimize etmek oldukça karışıkır. Ancak yaygın optimizasyon yöntemleri kullanarak, oyma dişli matkaplar için en iyi parametre kombinasyonlarının seçilmesiyle en düşük maliyeti oluşturan matematiksel modeller türetilmiştir. Bu matematiksel modellerden en kapsamlı ve en yaygın olanı Bourgoyne ve Young (BY) yöntemidir. BY yönteminde en iyi ilerleme hızını tahmin edebilmek için sekiz parametre içeren en az otuz girdi veri setine ihtiyaç vardır. Bu otuz veri seti, ya bir sahadaki otuz farklı kuyudan ayrı ayrı şeyl zonlarından alınmış olmalı ya da bir kuyuda otuz farklı derinlikteki şeyl noktalarından elde edilmiş olmalıdır. Herhangi bir sebeple elde yeterli veri olmadığı durumlarda BvY yönteminin doğruluğu azalmakta ve önemli hatalara yol açmaktadır. Bu nedenle, verinin yetersiz olduğu durumlarda alternatif yöntemler kullanılması zorunludur. Bu alternatif yöntemlerden en yaygın olanı, yapay öğrenme yöntemlerinin en etkililerinden biri olan Destek Vektör Regresyonu (DVR)'dur. DVR'nin ilerleme hızı tahmini problemine uygulanabilirliği literatürde ilk kez bu çalışma ile gösterilecektir.

Bu çalışmada, ilerleme hızını tahmin etmek için iki farklı regresyon tekniği kullanılmıştır. İlk teknik, BvY'nin uyguladığı çoklu regresyon analizidir. İlerleme hızı probleminde bir bağımlı parametreyi tek bir bağımsız parametre ile bağdaştırmak mümkün değildir. Ayrıca, bu parametreler aynı zamanda birbirlerini de etkilemektedir. Bu nedenle, bu şekilde birden çok parametrenin bulunduğu

durumlarda tekli regresyon analizi yapmak mümkün olmamaktadır. Çoklu regresyon analizi, parametrelerin birbirleri ile olan ilişkilerini çeşitli yöntemlerle belirleyip, her bir parametre için korelasyon katsayısı hesaplar. Daha sonra bu korelasyon katsayıları sayesinde tahmin modeli oluşturulur. Tahmin modeli oluşturulduktan sonra uygunluk katsayısı adı verilen R^2 değeri hesaplanır. Bu sayede katsayıların geçerliliği ve modelin uygunluğu gözlemlenir. R^2 değeri 1'e ne kadar yakınsa oluşturulan model o kadar geçerlidir. Çalışmada irdelenmiş olan ilerleme hızı probleminde birbiriyle ilişkili sekiz parametre bulunmaktadır. Her bir sekiz parametre için ise otuz adet veri seti vardır. Bu otuz veri seti, en küçük kareler yöntemi kullanılarak modellenir ve sekiz adet korelasyon katsayısı bulunur.

İkinci teknik ise güçlü bir yapay öğrenme yöntemi olan Destek Vektör Makinesi (DVM)'nin regresyon modelidir. Büyük miktarlardaki verilerin elle işlenmesi ve analizinin yapılması mümkün değildir. Bu tür problemlere çözüm olması amacıyla yapay öğrenme (makine öğrenmesi) yöntemleri geliştirilmiştir. Bu yöntemler, eldeki (geçmiş) verileri kullanarak, bu verilere en uygun modeli bulmaya çalışırlar. Bu işleme, öğrenme işlemi adı verilir. Model oluşturulduktan sonra yeni gelen (gelecek) veriler, bu modele göre analiz edilip sonuç üretilir.

Yapay öğrenme yöntemleri farklı uygulamalara, analizlere ve beklentilere göre gruplara ayrılır. Bu gruplardan en yaygın olanları sınıflandırma, kümeleme ve regresyondur. DVM, sınıflandırma konusunda kullanılan oldukça etkili ve basit yöntemlerden birisidir. DVM'de sınıflandırma işlemi için aynı düzlemde bulunan iki grup, aralarına bir sınır çekilerek birbirinden ayrılır. Sınırın çekileceği yer ise iki grubun da elemanlarına en uzak olan yer olması gerekmektedir. Bu işlem, iki gruba da yakın ve birbirine paralel iki sınır çizgisi çekilerek yapılır. Daha sonra bu sınır çizgileri birbirlerine yaklaştırılarak ortak sınır çizgisi üretilir. DVM'de sınıflandırma işlemi iki grup arasında yapılabileceği gibi ikiden çok grup arasında da yapılabilir.

DVM'de regresyon ile sınıflandırma arasında matematiksel olarak çok fark bulunmamaktadır. İki yöntem de yapısal risk minimizasyonu ve istatistiksel öğrenme teorisi ile çalışır. Çıktı olarak sınıflandırma bir çeşit etiket (label) verirken, regresyon bir sayı verir. Bu çalışmada, tahmin edilmesi istenen değer ilerleme hızı, yani sayısal bir değer olduğu için DVM'nin regresyon modeli kullanılmıştır. Bu model Destek Vektör Regresyonu (DVR) olarak adlandırılır. Yöntem, kullanılmak istenen veri setinin öğrenme (train) ve test olmak üzere iki alt veri setlerine bölünmesi ile uygulanır. Öğrenme veri seti kullanılarak, ilgili parametreler ve gözlemler arasındaki ilişki belirlenerek bir regresyon modeli oluşturulur. Daha sonra test veri seti, oluşturulan bu model üzerine uygulanarak hedef değer tahmin edilir.

Bu çalışmada DVR'nin yaygın modellerinden biri olan Epsilon-duyarsız kayıp fonksiyonu ve nü kontrol parametrelili model kullanılmıştır. Bu modelde, öğrenme veri setindeki her bir gözlem değerinden en fazla Epsilon kadar sapma yapacak ve mümkün olduğunca düz olacak şekilde bir fonksiyon bulunur. Diğer bir deyişle, Epsilon'dan küçük olan hatalar göz ardı edilir; fakat Epsilon'dan büyük sapmalar kabul edilmez. Belirlenen epsilon bandı civarında da gevşek değerlerin tolare edilebilirliğini belirleyen bir penaltı parametresi belirlenir. ν parametresi ise epsilon bandını kontrol edebilen kullanıcı tarafından belirlenen bir parametredir.

Bu tez çalışmasında DVR yöntemi kullanılarak ilerleme hızı tahminleri BvY çalışmasındaki veri seti kullanımıyla gerçekleştirilmiştir. Veriler sekiz parametre kapsamında tanımlandığı için genel bir yaklaşım olarak parametre sayısının iki katı olan 16 veri, üç katı olan 24 veri ve tamamı girdi verisi olarak kullanılmıştır. Geri

kalan veriler ise test seti olarak kullanılmıştır. Veri seçimi rastgele olabildiği için çok sayıda analiz kombinasyonu ortaya çıkmaktadır. Veri seçiminde tercih edilen yaklaşıma bağlı olarak DVR yönteminin uygulanması farklı durumlar ve senaryolar için incelenmiştir.

Farklı yöntemler ile elde edilen sonuçlar, her bir senaryo için ait olduğu durum altında irdelenmiştir. Sonuçlarda en iyi tahmin yapan yöntemin seçilen veri setine bağlı olduğu görülmüştür. Öğrenme veri setinin az olduğu durumlarda ilerleme hızı tahmini yapmak için DVR'nin çoklu regresyon yerine alternatif olarak kullanılabileceği belirlenmiştir.

Makine öğrenmesi yöntemlerinin en etkililerinden biri olan DVR, ilerleme hızı optimizasyonu için literatürde ilk kez bu tez çalışmasında kullanılmıştır. Böylelikle, DVR'nin ilerleme hızı optimizasyonu problemine genelleştirilmiş bir çözüm sunabileceğinin araştırılması gibi yeni araştırma alanı ortaya çıkmıştır.

1. INTRODUCTION

Nowadays, exploration of petroleum, natural gas and geothermal sources is gaining importance due to the unending increase of demand on the energy sources. Drilling of oil and gas wells, therefore, has gained remarkable technological improvements during recent years. There is an ongoing resource exploration processes beneath the unexplored subsurface and in the subsea deeper than 3000 meters of water depth in order to fulfill the needs. Because of the cost-efficient policy of oil companies, importance of lowering cost, increasing performance and reducing problems have risen. Now, these vital issues can be optimized better by employing advancing technology and widely used computer science.

Optimization is an important tool in most of the engineering problems. Optimization can be defined as “the process of finding the conditions that give the maximum or minimum value of a function.” (Rao, 2009, p. 1). In Figure 1.1, the corresponding minimum and maximum values of a function can be seen as an example.

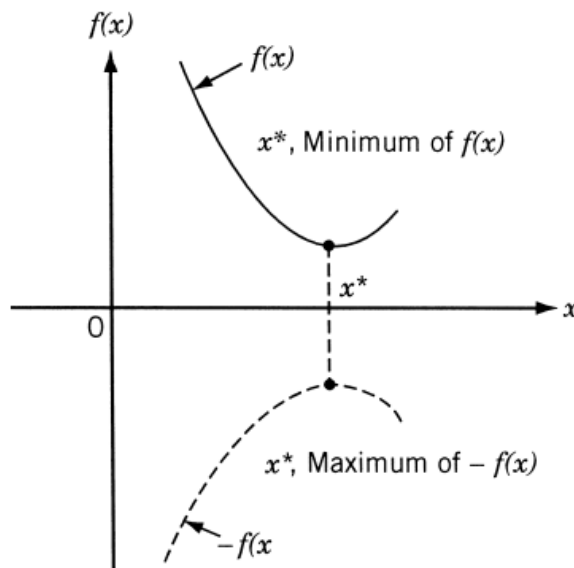


Figure 1.1 : Minimum and maximum of a function (Rao, 2009, p. 2).

To use optimization, four concepts should be described: Objective, a perceptible assessment of the system performance; variables (unknowns), the certain characteristics of the system that the objective depends on; constraints, the restriction

of the variables; and modeling, the determination of objective, variables and constraints for a specific problem (Nocedal and Wright, 1999, p. 1).

Once the modeling procedure is completed, in other words, the problem is mathematically formulated, numerical and computational techniques are used to find the optimum solution for the given problem in the reference of mathematical programming framework, as a part of operations research, which brings generic and flexible approaches in formulation and solution of engineering optimization problems (Iqbal, 2013, p. 10). Different methods of operations research is listed in Table 1.1.

Table 1.1 : Methods of operations research, adapted from (Rao, 2009, p. 3).

Mathematical Programming or Optimization Techniques	Stochastic Process Techniques	Statistical Methods
Calculus methods	Statistical decision theory	Regression analysis
Calculus of variations	Markov processes	Cluster analysis
Nonlinear programming	Queueing theory	Pattern recognition
Geometric programming	Renewal theory	Design of experiments
Quadratic programming	Simulation methods	Discriminate analysis
Linear programming	Reliability theory	
Dynamic programming		
Integer programming		
Stochastic programming		
Separable programming		
Multiobjective programming		
Network methods		
Game theory		
<u>Modern or Nontraditional Optimization Techniques</u>		
Genetic algorithms		
Simulated annealing		
Ant colony optimization		
Particle swarm optimization		
Neural networks		
Fuzzy optimization		

Cost efficiency in exploration operations has always been primary optimization problem for oil companies. Moreover, drilling operations constitute the major part of these exploration costs. The main aim in cost control while drilling is to minimize total well costs (Moore, 1986, p. 17). However, sustainability of cost control is another problem that can be faced during a drilling operation. For example, experiencing one good result of a particular problem may give contrary result for a similar problem. Thus, the logic of encountering cost efficiency problems might differ individually among the companies.

Drill bits are essential instruments to drill a hole in the earth's surface. Drill bits are among the primary drilling equipment, and selection of the optimum bit and operating conditions is one of the main problems that can be faced during a drilling operation (Bourgoyne Jr. et al, 1991, p. 190). In addition, drill bits are the primary part needs to be changed frequently due to its quick wearing nature. In order to reduce the drilling cost, drill bits are needed to operate well in long time. Therefore, the condition of drill bits should be monitored carefully either before, after or during the drilling operations.

The aim of assessing the condition of a dull bit is for making out which time interval the bit should be used in (Bourgoyne Jr. et al, 1991, p. 214). There are two possible situations for a dull bit condition. Firstly, the bit should have been pulled out earlier from the well. This situation is called as “pull one green” (Langenkamp, 1994, p. 337). It means if a drill bit is pulled green, the bit still has a remaining lifetime that can be used again. If this situation happens, there can be an occurrence of high rig cost by wasting time on unnecessary tripping (Bourgoyne Jr. et al, 1991, p. 214). Alternatively, as another result, when the bit that pulled green is decided to be used in another well by looking the unworn physical condition, it can fail in a quite short time and create unwanted costs. As a result, it is very crucial to determine the optimum bit changing time when considering high drilling rig cost, particularly in offshore drilling.

The second situation provides a basis to the bit optimization –or drilling optimization. A drill bit will going to wear while drilling as time goes by (Devereux, 2012, p. 147). Correspondingly, the rate of penetration (ROP) rapidly decreases as bit wears in time (Adams, 1985, p. 206). As a result, this condition creates extra cost. The infographic shown on the Figure 1.2, as a similar approach to the concept shown

in Figure 1.1, provides an example of non-linear relationship between time and drilling cost, which includes an optimum (minimum) value.

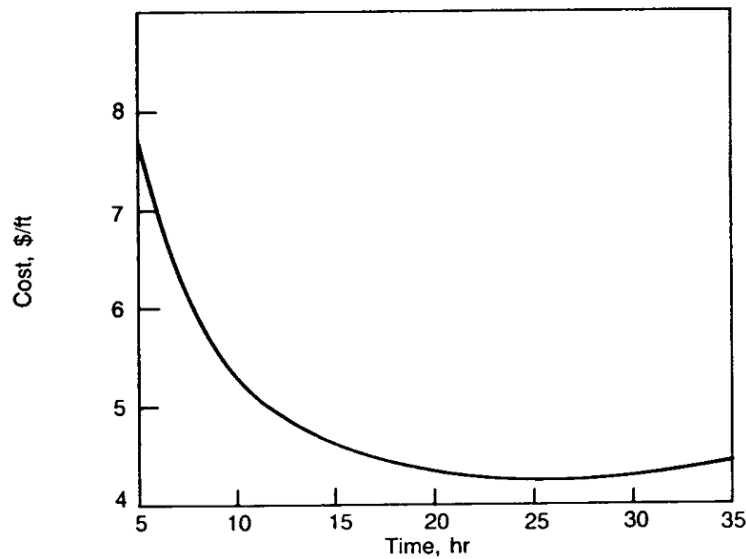


Figure 1.2 : An example of time and cost relationship (Moore, 1986, p. 28).

Theoretically stating, terminating the bit run is the optimum point where the drilling cost is the minimum. However, there is no commonly accepted rule of thumb among the operators on defining this optimum time since there are also other variables (difficult to be determined) affecting the cost-time curve. In addition, couple hours before and after the point of minimum cost-time curve may not alter the overall drilling cost much. The other parameters affecting the ROP and drilling cost except the terminating bit time and the bit wear are weight on bit (WOB), drill string rotation (rpm), drillability of formations, depth, mud density, formation pressure, drilling fluid hydraulics, jet impact force, bearing condition of bit, etc. It can be stated that maximizing ROP and/or minimizing drilling cost per footage drilled is a multi-parameter optimization problem requiring the usage of different optimization techniques.

In addition, if the bit is used in a long time interval, it may be shattered and leave junk in the hole (Bourgoyne Jr. et al, 1991, p. 214). This situation will require workover operations in the well in order to continue drilling which means extra time loss and extra cost. Hence, to determine the safest time interval of bit use, the bit wear rate must be known at once (Bourgoyne et al, 1991, p. 214).

In the past, several studies had been made to determine the optimum bit life by the means of optimum ROP and minimum cost. Yet, the effect of variables that effects

ROP is not fully recognized and complicated to model (Bourgoyne Jr. et al, p. 232). So, several mathematical models have been developed to optimize ROP and varying drilling parameters. The most commonly used model is introduced by Bourgoyne Jr. and Young Jr. (1974), which is deeply investigated in this study.

Using current mathematical models manually requires significant calculation time. This situation can cause severe time loss and high possibility of calculation errors. In addition, it is impossible to analyse and optimize huge amount of datasets on paper. As mentioned before, thanks to the improving technology, computer softwares are very helpful to solve these kind of optimization problems. In the last years, there is a concept named “Machine Learning” developed for learning from data to make future predictions. According to Alpaydm (2010), “Machine learning is programming computers to optimize a performance criterion using example data or past experience.” In other words, machine learning is a useful conception when a dataset is available to analyse in some way. In this study, regression approach of Support Vector Machines (SVM), or Support Vector Regression (SVR), which is one of the most effective machine learning methods is used.

This thesis is organized as follow:

- Chapter 2 contains a general literature review on methods developed for optimizing ROP. Also, several new algorithms that applied to ROP optimization problem is to be explained.
- Chapter 3 is about the statement of the problem and the scope of this thesis study.
- Chapter 4 represents the Multiple Regression (MR) method. In this chapter, theory of the MR analysis, different approaches and the application to the ROP optimization problem, which is the study of Bourgoyne Jr. And Young Jr. (1974) is explained in details.
- Chapter 5 is the theoretical explanation of SVR method. First the SVM method and its different approaches are described. Also, this is the algorithmic part of the thesis. Thus, the softwares and toolboxes for compiling and running the algorithms is going to be clarified. Then, the optimization and tuning process of SVR for the problem is explained.

- Chapter 6 is the part that the results are given. The results of two methods, MR and SVR, for numerous cases are going to be revealed, evaluated and compared to each other.
- Chapter 7 is finally about the conclusion and recommendations.

Chapters 2-3 are only an introduction to the topic. Chapters 4-5 are the theoretical and technical parts that deeply explains the mathematical concepts of MR and SVR methods. Also, Chapter 5 contains an example syntax of the code that performs MR and SVR, for those who wish to study the algorithm.

2. LITERATURE REVIEW

In past years, various studies have been made to optimize rate of penetration (ROP) and maintain cost control of drilling operations. Basically, it is desired to achieve the minimum cost and optimizing drilling parameters as well as obtaining the maximum rate of penetration (Bahari and Seyed, 2009, p. 451). The research field of the optimization of drilling operations will maintain its importance due to high oil prices in general.

There are different approaches and methods to sustain optimal drilling conditions such as an analytical method of Galle and Woods (1963), a Monte Carlo approach (Reed, 1972), and a numerical method of Bourgoyne Jr. and Young Jr. (1974), which provides a basis for this thesis study.

On the other hand, new computerized technologies have been started to be used in drilling optimization problems in recent decade like Trust-Region Approach (Bahari and Baradaran Seyed, 2007), Genetic Algorithms (Bahari et al, 2008), a real time multiple regression based method (Eren and Ozbayoglu, 2010), Artificial Neural Networks (ANN) (Bataee and Mohseni, 2011), and a progressive stochastic method (Rahimzadeh et al, 2011). Moreover, the research of practicability of SVM method in ROP optimization problem is performed first time in literature with this study.

2.1 Former Studies on ROP Optimization

In this part, the studies done before 2000 are reviewed.

First drilling parameters optimization study has been done by Speer (1958). In his study, an elementary method to obtain the best combination of weight on bit (WOB), rotary speed, and hydraulic forces that yields minimum cost was handled. Experimental relations were established to determine the effect of WOB, rotary speed, and hydraulic forces on ROP. It was showed that formation drillability is important factor for optimum weight. Similarly, WOB is the decisive parameter for

optimum rotary speed. These relationships produced a chart for determining optimum drilling conditions from a limited field data.

Garnier and van Lingen (1959) accomplished laboratory drilling experiments by using drag and roller cone bits at elevated mud, pore and confining pressures on rocks differing in strength and permeability. They found that the pressure difference between mud and pore pressure is the main factor that reduces ROP at a certain depth.

Graham and Muench (1959) introduced optimum combinations of rotary speed and WOB in order to minimize the cost. In their study, a mathematical analysis was performed to find out if optimum combinations of bit weight and rotary speed are present that minimizes the cost to drill specific depth intervals. They used a field data to perform the analysis and derive empirical correlations in terms of bit life expectancy.

Maurer (1962) derived an ROP formula for roller cone bits from their rock cratering mechanisms (2.1). He found that the ROP is directly proportional to rotary speed and to bit weight, and inversely proportional to bit diameter and rock strength. The mentioned formula is,

$$R = \frac{4}{\pi d_b^2} \frac{dV}{dt} \quad (2.1)$$

where R is the ROP, d_b is the bit diameter, V is the volume of the rock removal, and t is the time.

Galle and Woods (1963) presented a study of the best constant bit weight and rotary speed for rolling cutter bits to obtain minimum cost. They performed optimization to determine the best combination of constant weight and rotary speed, the best weight for any rotary speed, and the best rotary speed for any weight. Additionally, they introduced an equation of ROP as,

$$R = C_f \frac{\bar{W}_b^k r}{a^p} \quad (2.2)$$

where,

$$\bar{W}_b = \frac{7.88W_b}{d_b} \quad (2.3)$$

and,

$$r = \left[e^{-100/N^2} N^{0.428} + 0.2N(1 - e^{-100/N^2}) \right] \text{ (for hard formations)} \quad (2.4a)$$

$$r = \left[e^{-100/N^2} N^{0.75} + 0.5N(1 - e^{-100/N^2}) \right] \text{ (for soft formations)} \quad (2.4b)$$

and,

$$a = 0.928125h^2 + 6.0h + 1 \quad (2.5)$$

In these equations, R is the ROP, C_f is the formation drillability parameter, W_b is bit weight, d_b is hole or bit diameter, N is rotary speed, and h is bit tooth dullness. In addition, k 1.0 for most formations except very soft formations, 0.6 for very soft formations, and p is 0.5 for self-sharpening or shipping-type bit tooth wear.

Bingham (1965) stated an ROP equation based on laboratory data (2.6). In this equation, he assumed the threshold bit weight is negligible and ROP is a function of WOB and rotary speed.

$$R = K \left(\frac{W}{d_b} \right)^{a_5} \quad (2.6)$$

where R is the ROP, K is the drillability of formation, W is the WOB, d_b is the bit diameter and a_5 is the exponent that determined experimentally.

Eckel (1967) made microbit tests with a low permeability limestone. The tests were handled at constant bit weight and rotary speed with various properties such as flow rate and nozzle diameter. The tests showed that ROP with a constant circulation rate and nozzle velocity is a function of kinematic viscosity of the drilling fluid measured close to the bit nozzle shear rates, the effect of fluid properties and hydraulics on microbit ROP is defined by Reynolds number function, and ROP is independent of solid content and fluid loss at the same kinematic viscosity.

Wardlaw (1969) showed an elementary relationship between drilling efficiency and ROP in terms of rotary speed and WOB. The factors affecting these two parameters defined as differential pressure on bottom, mud characteristics, circulation rate, jet

velocity and rock-bit design. Moreover, it was indicated that drilling performance was affected with hydraulic horsepower.

Young Jr. (1969) developed and installed an on-site computer system to control bit weight and rotary speed. He indicated that the pre-actual tests showed that the system could be precious to reduce drilling costs. The ROP equation in terms of minimum cost drilling expressions was defined as,

$$R = \frac{K(W_b - M)N^\lambda}{1 + C_2H} \quad (2.7)$$

where K is formation drillability, W_b is bit weight, M is translation constant, N is rotary speed, λ is rotary speed exponent, C_2 is a constant, and H is normalized bit tooth height.

Lummus (1970) noticed in his study that mud and hydraulic forces are the main factors affecting drilling optimization. He also indicated that drilling operations could be optimized when the rig is efficient to provide enough hydraulic forces, necessary rotation, and WOB. He also made another study showing the emphasis of data obtaining for drilling optimization (Lummus, 1971). The study mainly focused on planning and data collecting. The important data requirements were expressed as computer inputs to calculate optimum controllable drilling variables.

Reed (1972) solved the variable weight-speed optimal drilling problem using Monte Carlo approach to create the drilling schedule that gives the minimum cost per foot. A fast computer program was written to perform linear and curvilinear smoothing techniques to create random and discrete paths in order to decrease the cost per foot at each step of the Monte Carlo iterations. He compared his results with Galle and Woods (1963) and indicated a performance improvement.

Study of Wilson and Bentsen (1972) were based on two variables, WOB and rotary speed, as well as the other parameters such as mud properties and bit type properly selected. They investigated to minimize the cost per foot drilled during a bit run, for a selected interval and over a series of intervals. It was also indicated that each of the investigations showed remarkable savings on costs. Additionally, the savings increased as the data requirements increased.

Bourgoyne Jr. and Young Jr. (1974) described a model (Bourgoyne and Young Model – BYM) based on a multiple regression analysis of a drilling data acquired in short time intervals. The study involved the effect of formation strength, formation depth, formation compaction, pressure differential across the bottom hole, bit diameter, bit weight, rotary speed, bit wear and bit hydraulics. The study presented involves in calculation steps for selecting bit weight, rotary speed, and bit hydraulics and formation pressure with applying regressed model on the drilling data. This model is the key point of this thesis study and is going to be explained in details in Chapter 4.

Tansev (1975) presented a heuristic approach involving the interaction of raw data, regression and optimization techniques. In his study, ROP and bit life predictions has been made for several bit runs with regression analysis. He accounted for three variables, namely WOB, rotary speed and hydraulic. The ROP equations intersected with drilling cost equations and determined optimum variable values for minimum cost. He also mentioned that, drilling a uniform formation and using similar bits were the best choices for his approach.

Doiron and Deane (1982) performed laboratory tests to measure the effects of hydraulic cleaning parameters on ROP for soft formation insert bits. It is mentioned that the ROP is directly proportional to bit hydraulic horsepower and reversely proportional to nozzle size or number. Moreover, decrease in drilling cost caused by the measured response of the ROP to enhance bottom hole cleaning was compared to operating costs required to fulfill the additional hydraulic power. It is reported that excessive bit hydraulic horsepower levels are cost effective for low ROP because of high overbalance pressure.

Bizanti and Blick (1986) made test exploring the variables affecting the cutting removals with dimensional analysis. They indicated the rate of cutting removals was a function of dimensionless parameters such as Reynolds number, rotational Reynolds number, and Froude number. At the end of a regression analysis, they presented R^2 and F-values to show the confidence level of the correlation. They also stated optimizing bottom hole cleaning would improve ROP in terms of minimum cost per foot.

Reza and Alcocer (1986) introduced a new non-linear multidimensional, dimensionless mathematical drilling model for deep well operations using the Buckingham Pi theorem dimensional analysis. They used WOB, rotary speed, kinematic viscosity, volumetric flow rate, bit diameter, differential pressure, temperature, and heat transfer coefficient as controlling parameters. The ROP model is given in (2.8).

$$\frac{R}{Nd_{be}} = C_1 \left(\frac{Nd_{be}^2}{v_k} \right)^{e_1} \left(\frac{Nd_{be}^3}{Q} \right)^{e_2} \left(\frac{Ed_{be}}{W} \right)^{e_3} \left(\frac{pd_{be}}{W} \right)^{e_4} \quad (2.8)$$

In (2.8), R is ROP, N is rotary speed, d_{be} is bearing diameter, v_k is kinematic viscosity, Q is volumetric flow rate, E is rock hardness, p is differential pressure, and W is WOB. In addition, C_1 , e_1 , e_2 , e_3 , and e_4 are the unknown exponents and proportionally constants, which were calculated by the statistical regression curve fitting method.

Simmons (1986) epitomized a technique including numerous drilling parameters such as hydraulics, WOB, and rotation to get maximum drilling efficiency, which can be controlled real-time on the field. The technique also allows the drilling supervisor on location to fine tune the parameters by modifying it to achieve the optimum drilling performance. This study also counted as one of the milestones of real-time drilling optimization studies (Eren, 2010, p. 30).

Warren (1987) stated that the ROP gathered from roller-cone bits are limited because of the cuttings occurrence and cuttings removal. He presented an ROP model, which contains the effect of initial chip formation and cuttings removal (2.9). He also indicated that due to local cratering and global cleaning effects there could be a decrease in ROP at high borehole pressures. The mentioned ROP model was presented as,

$$R = \left(\frac{c_1 E^2 d_b^3}{NW^2} + \frac{c_2}{Nd_b} + \frac{c_3 d_b \gamma_f \mu}{F_{jm}} \right)^{-1} \quad (2.9)$$

where R is ROP, c_1 , c_2 , c_3 are dimensionless constants, E is rock hardness, d_b is bit diameter, N is rotary speed, W is WOB, γ_f is fluid specific gravity, μ is plastic viscosity, and F_{jm} is modified jet impact force.

Al-Betairi et al. (1988) reported an application of multiple regression analysis based on BYM. The model estimated optimum ROP, WOB, and rotary speed in terms of controllable and uncontrollable parameters. They used a field data taken from three wells in Arabian Gulf to validate the model. They indicated that the model coefficients are sensitive to the number of data points.

Lubinski (1988) introduced three differential equations in terms of ROP, rate of tooth wear, and rate of bearing wear. It was stated that if the basic conditions were met, these equations could be used to optimize the drilling process. The mentioned basic conditions are acceptable bottom hole cleaning, properly selected rock-bit, and homogeneous formation.

Maidla and Ohara (1991) developed a computer software that minimizes cost per foot for a single bit run and for the concurrent selection of a roller-cutter bit, bit wearing, WOB and drill string rotation. They compared their model (2.10) with BYM by using field data taken from five different offshore wells in a particular location. The outcome of the study indicates the ROP values of the fifth well can be predicted by using the coefficients of previous four wells with some cost saving. The model is described as,

$$R = \exp\left(b_1 + y_1 + \sum_{k=2}^6 b_k y_k\right) \quad (2.10)$$

where,

$$y_1 = \ln(Nd_b) \quad (2.11)$$

$$y_2 = \ln\left(\frac{W}{S_c d_b^2}\right) \quad (2.12)$$

$$y_3 = \frac{p_p - p_c}{S_c} \quad (2.13)$$

$$y_4 = 2 - 5 \times 10^{-5} \left(\frac{D}{d_b}\right) \quad (2.14)$$

$$y_5 = \ln\left(\frac{F_j}{S_c d_b^2}\right) \quad (2.15)$$

$$y_6 = -\frac{h}{d_b} \quad (2.16)$$

Here, R is ROP, N is rotary speed, d_b is bit diameter, W is WOB, S_c is compressive strength, p_p is pore pressure, p_c is circulating bottomhole pressure, D is well depth, F_j is jet impact force, and h is fractional tooth height worn away. Additionally, the constants b_1 through b_6 is calculated for each different bit choice.

Pessier and Fear (1992) handled full-scale simulator tests under laboratory conditions to develop and validate a specific energy in rotary drilling model, which is introduced by Teale (1965):

$$E_s = \frac{W}{A_B} + \frac{120\pi NT}{A_B R} \quad (2.17)$$

Here E_s is specific energy, W is WOB, A_B is borehole area, N is rotary speed, T is bit torque, and R is ROP. It was also reported that field data analysis gives good correlation results between simulator and field results.

Wojtanowicz and Kuru (1993) presented a leading approach to drilling optimization called dynamic drilling strategy. It is mentioned that dynamic drilling strategy is a new technique, which combines the theory of single-bit control and multi-bit program for a well. The method was compared to conventional methods and field practices in the simulation part. Furthermore, cost-saving potential of 25% and 60%, respectively, was estimated. Finally, it is stated that the method seems to be the best cost-effective approach for long-lasting Polycrystalline Diamond Compact (PDC) bits in terms of reducing the number of bits needed.

Rampersad et al. (1994) performed a utilization by creating a Geological Drilling Log (GDL) to optimize the drilling cost. The GDL is created by using specific drilling models for specific bit types for each formation profile interval taken foot by foot of the entire drilling section (2.18), (2.19).

The tricone bit model is defined as,

$$R = W_f \left[f_c(P_e) \left(\frac{aS^2d_b^3}{NW^2} + \frac{b}{Nd_b} \right) + \frac{c\rho\mu d_b}{F_{jm}} \right]^{-1} \quad (2.18)$$

The PDC bit model is defined as,

$$R = \frac{14.14N_sN(A_v - A_{vw})}{d_b} \frac{a_d}{N^{b_d}W^{c_d}} \quad (2.19)$$

Here R is ROP, W_f is wear function, $f_c(P_e)$ is chip holddown function, S is confined rock strength, d_b is bit diameter, N is rotary speed, W is WOB, ρ is mud density, μ is mud plastic viscosity, F_{jm} is modified jet impact force, N_s is number of diamond stress, A_v is the projected area of each diamond, A_{vw} is the projected area of worn section of a diamond, and a, b, c, a_d, b_d, c_d are constant coefficients.

Cooper (1995) builded up a graphical user interfaced software that allows a student or engineer to drill a well and optimize the drilling process. The software had three parts: A lithology editor which lets the student to create different layers having different characterization, mineralization, etc.; a settings editor which all the operational parameters can be defined at; and the simulator algorithm that calculates ROP and the rate of tooth wear as drilling goes by.

Mitchell (1995) introduced a contour method for determining optimal WOB and rotary speed for minimum cost per foot. The method involves structuring a plot like Figure 2.1 in two dimensional space.

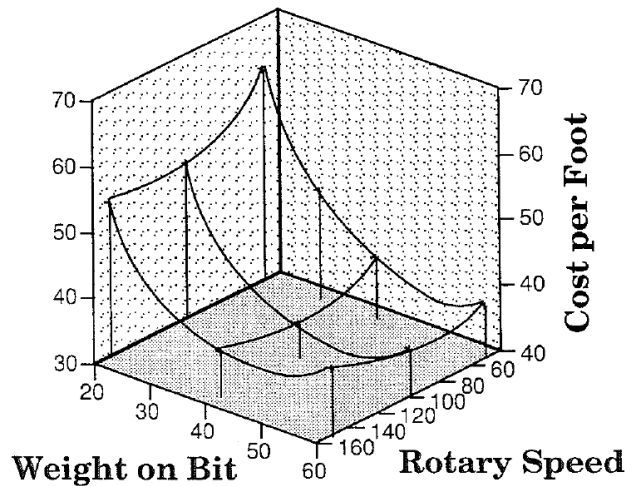


Figure 2.1 : Optimum WOB and rotary speed (Mitchell, 1995, p. 531).

This procedure can be done by plotting cost per foot contours with the past drilling data on the coordinates. The WOB is on the x-axis and the rotary speed is on the y-axis. Several bit runs are necessary to draw each line, which have either including or excluding values within the contour. The most cost effective bits have the nearest optimal WOB and rotary speed.

Fear (1996) created a method that classifies which parameters are controlling ROP in a specific group of bit runs. The method needs some input data such as mud logging unit data, geological information, and drill bit characteristics in order to estimate correlations between ROP and controlling parameters or other aspects. Finally, these correlations are used to make suggestions to maximize the ROP.

Barragan et al. (1997) developed a program for utilizing optimum drilling conditions using Monte Carlo Simulation and specially developed numerical methods. The output gives the optimum conditions that minimizes the cost per foot for each specific bit run through the entire well. The model does not rely on one particular ROP optimization method. The model itself tested with other methods such as BYM and Warren (1987).

Bengisu et al. (1997) introduced a new methodology to estimate ROP during drilling operations. The neural network approach analyzes acquired data to correlate several parameters such as bit type, WOB, depth, and rotary speed. They used two different data sets. First data set included 8000 measurements taken at specific wellbore conditions, and the second data set contained 500 measurements taken from several wells. They also used a rig floor simulator to acquire data. In addition, it is stated that the simulated data could provide additional understanding of different parameters such as formation abrasiveness, bit tooth wear, and bit bearing wear as a function of drilling time, which are not quite possible to measure during drilling.

Serpen et al. (1997) designed a project based on optimizing drilling processes to determine drilling parameters for decreasing time consumption and making predictions accurately. They created computer softwares using common drilling models available in the literature: Galle and Woods method, Constant-Energy Drilling method, drill-off tests, multiple regression approach, modified multiple regression approach, and drilling hydraulics optimization.

Dubinsky and Baecker (1998) stated the dynamic action of a drill bit is a crucial factor in drilling efficiency by affecting ROP, hole conditions and the tripping frequency. They developed a simulator to improve the perceptive of the dynamic behavior to get better results in drilling optimization. The simulator reproduces dynamic condition variables such as hook load, rotary speed, mud properties, flow rate, BHA, borehole parameters, etc. The system also simulates the major drilling

dynamic dysfunctions like bouncing effects, drill string vibration, torque shocks, etc. It is mentioned that the system has self-learning qualities depends on the amount of data used for training the model. Additionally, they referred their study as an efficient training toolbox for MWD operators and drilling staff.

Samuel and Miska (1998) presented an analytical model that optimizes performance and parameters for downhole motors (2.20). They also introduced a test called wear-off test, which connects an operating window. The model is described as below:

$$R = K \left(\frac{W}{4d_b} \right)^{e_1} \left(\frac{N}{100} \right)^{e_2} \quad (2.20)$$

Here R is ROP, K is formation drillability, W is WOB, d_b is bit diameter, N is rotary speed, and e_1 and e_2 are the exponents.

Millheim and Gaebler (1999) commenced a heuristic concept called Virtual Experience Simulation. They gathered data from 22 different wells and processed them by dividing into specific data sets in terms of geology, tripping, cementing, logging, ROP, and unscheduled events. By containing the field drilling experience and knowledge, these new data sets earns a value. It was believed by the authors that the study presents a new approach to evaluate field drilling experience.

2.2 Recent Studies on ROP Optimization

In this part, the researches made in and after 2000 are reviewed.

Akgun (2002) suggested a different approach in estimating drilling performance. This approach introduces a comparison of drilling performance in terms of technical limit of the ROP, which means a maximum attainable ROP without risking drilling safety. It is also stated that the technical ROP limit can be reached if these conditions are met: Selecting minimum mud weight that prevents any kick occurrence and borehole collapses, WOB and rotary speed selected at their possible maximum values, and optimum flow rate giving

John et al. (2002) described a modern information technology called InterACT, which aids data to be acquired from remote drilling wellsites in real-time. It is concluded that this technology provides cost efficiency in terms of rig time when

making operational decisions. The InterACT system is pictured as seen on Figure 2.2.

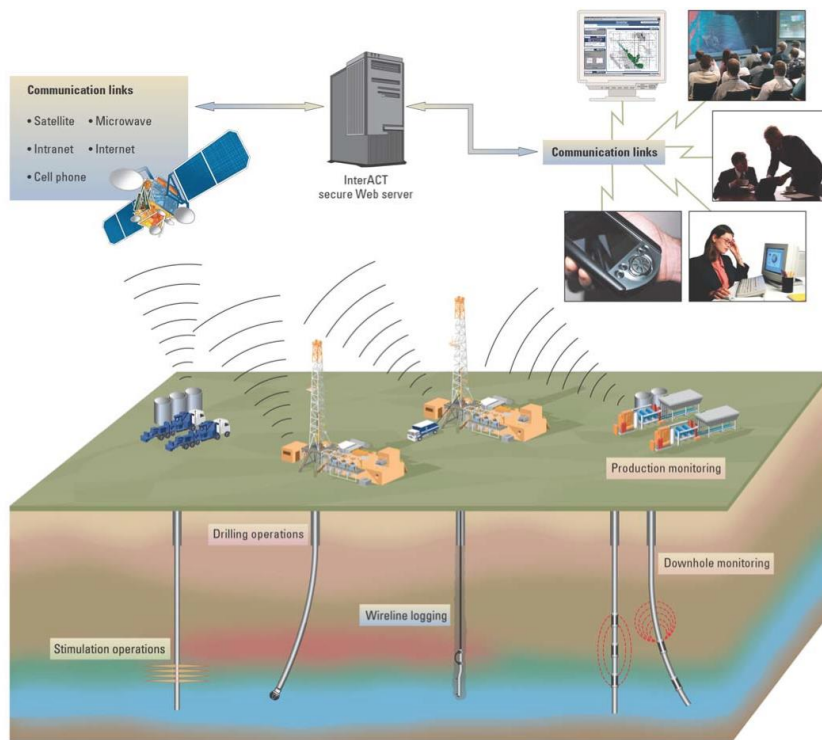


Figure 2.2 : The InterACT system (John et al, 2002, p. 2).

Gelfgat et al. (2003) published a book having a chapter about drilling optimization with a new approach used in the former Soviet Union. They stated their inspection and analysis of the data taken from an optimization process of US well forms on the statistical knowledge from previously drilled wells or specific bit runs. This data also allowed them to make adjustments on drilling parameters. They reported that the efficiency of the approach depends on the following criteria:

- A consistently high level of drilling technology and equipment provided by equipment and material supply companies,
- availability and smooth operation of instrumentation and recording equipment for drilling operations,
- exclusive use of the rotary drilling method that facilitates the optimization of drilling parameters within the acceptable range of rotational bit speed variation,

- a reliable system of payment for drilling crews that was not dependent on drilling results.

Ursem et al. (2003) demonstrated an implementation of a holistic approach called Real Time Operations Center with a collaboration of Shell and Halliburton by using high-end technology. The basis of the technology relies on learning from the past and making best decisions without the involving of people. Moreover, they presented the areas where people issues were taken under consideration, which are collaborative environment, real time monitoring, managing knowledge, change element and new initiatives, value challenges and investment versus scope.

Mochizuki et al. (2004) audited and classified technological elements in real-time optimization to analyze the value of real-time optimization projects. They analyzed case histories and demonstrated the effect on real-time optimization on the wellsite. They mentioned that the real-time optimization projects are quite difficult in type because of the involving of people, technology and process components.

Rommetveit et al. (2004) designed an innovative project for automating and simulating drilling operations, called Drilltronics. The system gathers all available surface and downhole drilling data in real-time to optimize the drilling process. If a sensor fails, an advanced mirroring model takes place to estimate the missing parameter. Furthermore, it is stated that to full perform the system, all surface and downhole drilling data must be acquired in real-time. Additionally, the drilling equipment must be computer controllable with an interface, which supports automatic response to the model's missing parameter calculations.

Caicedo et al. (2005) developed a new ROP prediction model based on the specific energy theory and confined compressive strength (2.21). This new model has been validated by field and laboratory tests for improving drilling performance, reducing well costs, and estimation of optimum drilling parameters. The model was symbolized as,

$$R = \frac{13.33\zeta N}{d_b \left(\frac{S}{e_m W} - \frac{1}{A_B} \right)} \quad (2.21)$$

Here R is ROP, ζ is bit-specific coefficient of sliding friction, N is rotary speed, d_b is bit diameter, S is confined rock strength, e_m is the mechanical efficiency, W is WOB, and A_B is borehole area.

Dupriest and Koederitz (2005) used the concept of Mechanical Specific Energy (MSE) to estimate the drilling efficiency of bits in laboratory conditions (2.22). It is indicated that the concept can be used more widely by drilling staff as a real-time optimization tool to maximize the ROP. They also mentioned the results exceeded the expectations by observing the average ROP of the six rigs selected for pilot was increased by 133%. The implementation of MSE can be formulated as,

$$MSE = \frac{480TN}{d_b^2 R} + \frac{4W}{\pi d_b^2} \quad (2.22)$$

Here MSE is mechanical specific energy, T is bit torque, N is rotary speed, d_b is bit diameter, R is ROP, and W is WOB.

Ipek et al. (2006) introduced a new method relating the performance of drill bits to cation exchange capacity of the drilled shaly formation. They built a relationship between drilling parameters by using normalized ROP, specific energy and cation exchange capacity (2.19), (2.23).

$$R_n = \frac{R}{\left(\frac{W}{d_b}\right)^2 \left(\frac{N}{60}\right)} \quad (2.23)$$

Here R_n is normalized ROP, R is ROP, W is WOB, d_b is bit diameter, and N is rotary speed. It is stated that the correlations indicated the potential diagnosis of bit balling and drilling ineffectiveness in over pressured shaly formation for PDC bits.

Iversen et al. (2006) presented an integrated system for monitoring and controlling drilling process. The essential part of the system is the models for fluid flow and drilling mechanics are updated sequentially in real-time. It is stated that unwanted occurrences can be detected quickly in comparison to calibrated models. They also optimized the calibrated models for staying in the safe limits for drilling operations. Finally, they reported that the system might be helpful to increase safety and reduce down time on drilling operations.

Milner et al. (2006) performed real-time data transfer from an offshore well to land support centers for drilling and production processes. The transmitted data was given quality by applicable personnel who have an availability of high speed broadcasting connection. They stated that by applying the real-time data transmission unwanted events and well shut-in operations were reduced in number.

Bahari and Baradaran Seyed (2007) wrote a computer software that applies four different mathematical methods to a field data to estimate the parameter coefficients. The methods were multiple regression, linear least square data fitting with non-negativity constraints, non-linear least square data fitting with Gauss-Newton, and non-linear least square data fitting with trust-region. The results were compared to each other. It is stated that the trust-region method is the best mathematical model to estimate BYM parameter coefficients. Additionally, it is reported that the trust-region method can be used in limited situations such as availability of a few data points and exceeding recommended parameter ranges.

Judzis et al. (2007) tested and identified limiting factors of ROP performance in extended depths, the advancement of circulation systems, and bit design criteria, which are all improving drilling performance. They performed sixteen tests for 6-inch bits at 10000 psi of wellbore pressure. The results showed that high-pressure tests with water and base oil gives excessive ROP values before mud circulation. It is also indicated, while drilling hard sandstone formation, ROP could drop about 75% if the mud is weighted.

Monden and Chia (2007) made a series of real case studies with an operation support center that inspects the range and scale of measurements in terms of drilling optimization and minimizing downtime. They improved ROP by advanced software modeling of drilling parameters. The cases included either the reduction of shock and vibration the drill string. It is mentioned that the ROP is directly improved by balancing weight on bit and rotary speed.

Osgouei (2007) proposed a modified BYM which includes the effect of formation compaction, formation pressure, equivalent circulating density, effective WOB, rotation of the bit, bit wear, hole cleaning, deviation, fluid loss, and hydraulics. In addition, he developed a bit wear model for roller cone and PDC bits. The model was tested with an offshore data taken from Persian Gulf. It was stated in comparison to

the field data, the model could predict the ROP with an error of $\pm 25\%$. He described the regression variables as,

$$y = \ln(R) \quad (2.24)$$

$$x_2 = 8800 - TVD \quad (2.25)$$

$$x_3 = TVD^{0.69}(g_p - 9) \quad (2.26)$$

$$x_4 = TVD(g_p - \rho_c) \quad (2.27)$$

$$x_5 = \ln\left(\frac{W/d_b}{4}\right) \quad (2.28)$$

$$x_6 = \ln\left(\frac{N}{100}\right) \quad (2.29)$$

$$x_7 = -h \quad (2.30)$$

$$x_8 = \ln\left(\frac{F_j}{1000}\right) \quad (2.31)$$

$$x_9 = \ln\left(\frac{A_{bed}/A_{well}}{0.2}\right) \text{ (for roller - cone bits)} \quad (2.32a)$$

$$x_9 = \ln\left(\frac{A_{bed}/A_{well}}{0.35}\right) \text{ (for PDC bits)} \quad (2.32b)$$

$$x_{10} = \ln\left(\frac{v_{actual}}{v_{critical}}\right) \quad (2.33)$$

$$x_{11} = \ln\left(\frac{C_c}{100}\right) \text{ (for roller - cone bits)} \quad (2.34a)$$

$$x_{11} = \ln\left(\frac{C_c}{25}\right) \text{ (for PDC bits)} \quad (2.34b)$$

Here R is ROP, TVD is total vertical depth, g_p is pore pressure gradient, ρ_c is equivalent circulating density, W is WOB, d_b is bit diameter, N is rotary speed, h is

final bit tooth dullness, F_j is jet impact force, A_{bed} is area of cuttings bed, A_{well} is area of the well, v_{actual} is mud velocity in annulus, $v_{critical}$ is mud critical velocity in annulus, and C_c is cuttings concentration for a stationary bed (by volume), corrected for viscosity.

Remmert et al. (2007) implemented an approach to manage ROP in real-time by customizing surveillance items to maximize bit cutter efficiency and energy transmission thru bit, simultaneously. They redesigned the hydraulic horsepower to decrease the effect of cuttings and monitored MSE regularly to optimize WOB and rotary speed in order to reduce vibrations and energy loss.

Strathman et al. (2007) built a system that captures data from the well in real-time, stores in a historian, and makes available in trends in time. The system could capture up to 200 parameters per well with the frequency of 5 seconds through 20 wells. The solution was adjustable by the operator using different data and time sequences. It is also mentioned that the system makes companies reduce their cost to access the well data by letting them develop better plans.

Tollefsen et al. (2007) presented a study performed in the Gulf of Mexico that permits an operator to drill safely in a very limited hydraulic envelope and eliminate a string of casing by using Logging-While-Drilling (LWD) tools. In this study, a predrill model created by using the velocities exported from a 3D modelling of northern Gulf of Mexico derived from checkshots and sonic logs. Furthermore, a new method was found to optimize drilling performance in a particular well by the help of LWD tools. The driller could maintain the wellbore stability via simultaneously updated LWD annular pressure measurements. In addition, it is stated that a full understanding of the hydraulic forces on a borehole can improve the ROP, safety, casing string design, kick prevention and optimized well completions.

Iversen et al. (2007) introduced a new drilling control system installed on a rig that improves real-time optimization and automation. A field test was performed to identify the incorporation of calibrated models in terms of efficiency, safety and reliability. The system uses calibrated dynamic process models to calculate operational parameters. They stated that the automation system provides improved control of well conditions.

Bahari et al. (2008) implemented Genetic Algorithm (GA) to BYM to determine parameter coefficients with high accuracy. It is stated that GA can make improvements on coefficient estimation in situations, which BYM leads physically meaningless values. They also tested their new method with field data. It is reported that simulation results validated the new method on coefficient estimation.

Iqbal (2008) made a simple analysis that re-evaluates the optimum values for the parameters like rotary speed and WOB by applying Galle and Woods (1963) method. He created a simple program that can change the specifications of optimum values and estimates the maximum ROP within available restrictions. Thus, the drilling cost decreases due to the increase of the ROP.

Bahari and Baradaran Seyed (2009) investigated a case study in the Iranian Khangiran gas field. They performed a combination of two optimization techniques, which are comparative methods (cost per foot and specific energy) and mathematical methods. The summary of the optimization algorithm is shown in Figure 2.3.

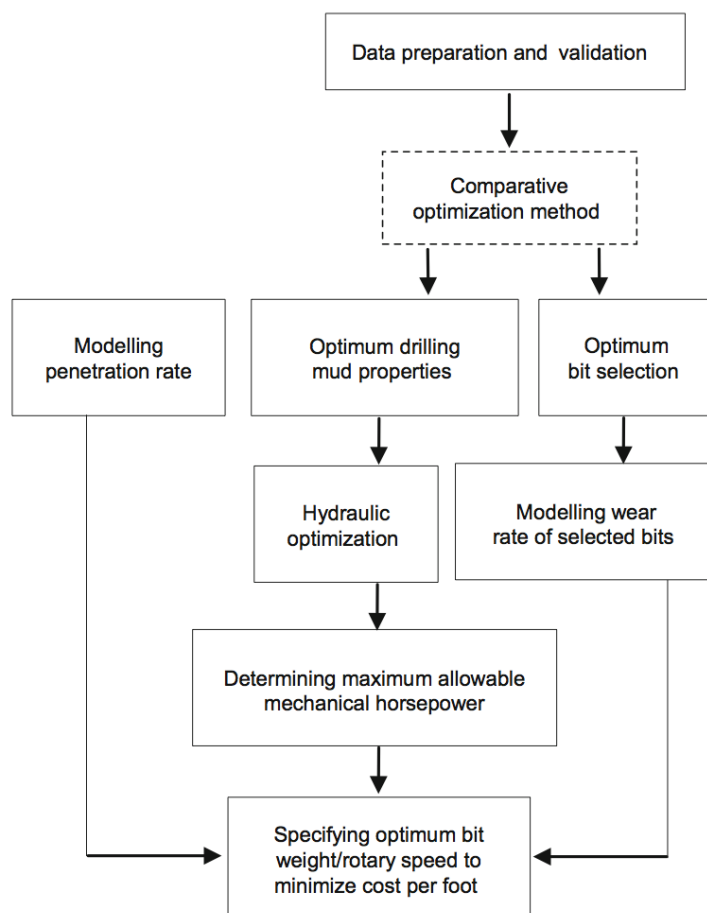


Figure 2.3 : Algorithm of the mathematical optimization procedure (Bahari and Baradaran Seyed, 2009, p. 452).

As a result of the study, it is stated that systematic combination of theory and application provides remarkable saving in time and money for drilling operations.

Cayeux et al. (2009) designed a system based on an application of real-time process models to calculate hydraulic and mechanical forces, which requires a huge set of input data. Results of the calculations were applied directly to drilling control system that affects pipe acceleration, velocity, and the pump startup profile. The most important concept of the system is the data to be correct and high quality.

Aadnøy (2010) indicated that the flow rate and the ROP must remain in specified limits to assure good hole cleaning performance. An example of presented charts is shown in Figure 2.4.

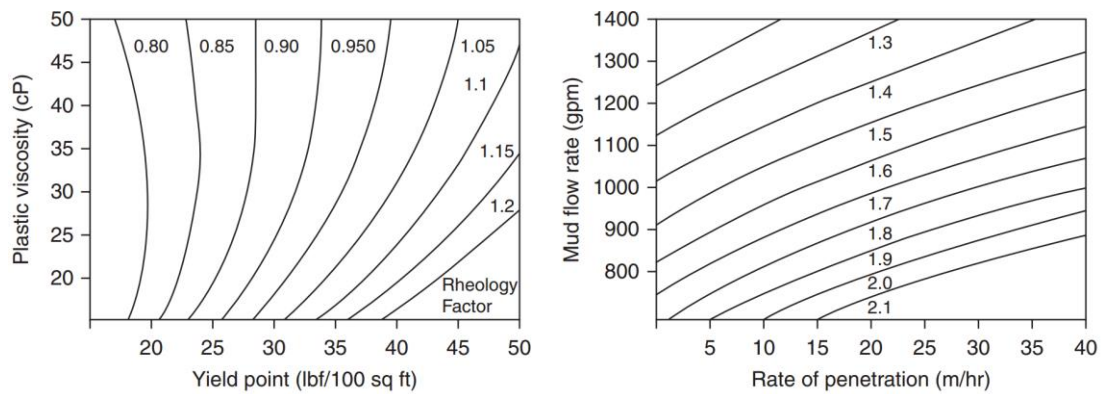


Figure 2.4 : Hole cleaning chart example (Aadnøy, 2010, p. 26).

Eren and Ozbayoglu (2010) developed a model based on BYM for drilling optimization using data gathered from modern well monitoring and data recording systems. The model was designed for predicting ROP as a function of valid drilling parameters. The general equation of ROP is optimized for effective functions at each individual data point. They also built a computer network system to pipe data simultaneously from the rig. They also mentioned that the ROP could be optimized in real-time as a function of independent variables such as weight on bit, rotation speed, mud weight, and formation characteristics.

Hareland et al. (2010) presented an experimental setup for rock failure caused by the inserts of roller cone bits by processing experimental data. They developed a new ROP model derived directly related to the fractures exerted by a single insert (2.35). Moreover, the study involved prediction of ROP and rock compressive strength with an offset data. It is indicated that verification of the model is done by stating good

estimation of the interaction between the bit and the bottom hole rock. The ROP predictions evaluated from the model matched the data taken to be analysed. The new model is described as,

$$R = K \frac{80n_i m N}{d_b^2 \tan^2 \psi} \left(\frac{W}{100n_i \sigma_p} \right) \quad (2.35)$$

Here R is ROP, n_i is the number of insert in contact with the rock at the bottom, m is the number of insert penetrations per revolution, N is rotary speed, d_b is bit diameter, ψ is chip formation angle, W is WOB, and σ_p is the ultimate strength of rock at a differential pressure.

Moradi et al. (2010) introduced a soft computing approach for ROP prediction. They applied a method based on K-mean clustering to a field data to show that the accuracy of their model is higher than the other statistical models like Bourgoyne Jr. and Young Jr. (1974). The flowchart of the study is demonstrated on Figure 2.5.

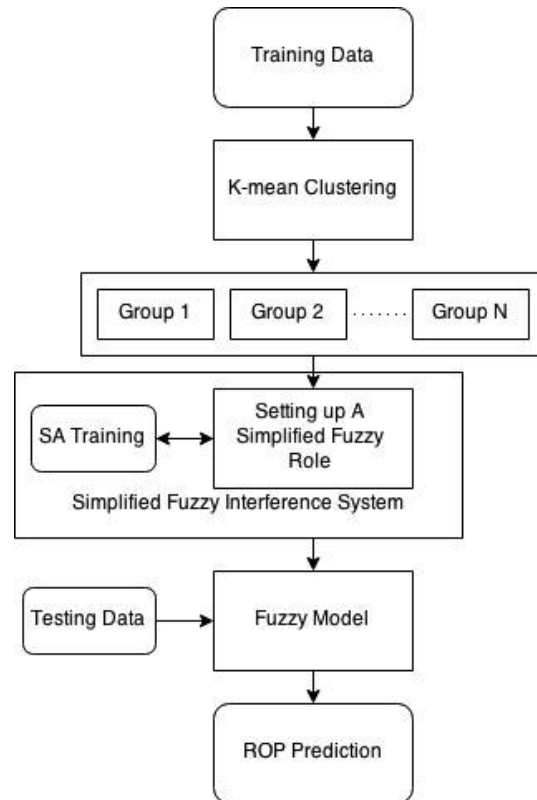


Figure 2.5 : Flowchart of a soft computing approach, adapted from (Moradi et al, 2010, p. 1585).

Wu et al. (2010) discussed three main factors affecting ROP in their study. The factors are rock types, insert types, and insert wear. To analyse the effect of these factors on ROP, they developed a simulator. It was found as a result of simulation that the decrease in ROP for a limestone formation is greater than that for shale as the bit wears. However, if the WOB is very small the results showed the opposite results. It was also indicated that when WOB is low the effect of dull grade on ROP is differ in rock types.

Bataee and Mohseni (2011) developed an optimization program that can be used in optimizing drilling parameters and estimating drilling time properly. They used ANN in their optimization program. It is concluded that ANN can find the actual values of parameters that maximize ROP despite no equation is capable of doing such estimation. The results also show that mud weight and ROP are inversely proportional. The best ROP was selected by achieving previous ROP using modelled function and applying proper drilling parameters.

Esmaeili et al. (2012) conducted experimental tests with a laboratory scaled drilling rig to explore the effect of drill string vibration on ROP optimization. The sensors recorded vibration data and other drilling parameters. In the experiments, artificial sandstone cubes were drilled with several ranges of WOB and rotary speed. Linear and non-linear models were developed to analyse the data with the help of ANN for given data acquiring frequency. It is stated that the modelling has an important effect on ROP optimization.

Gidh et al. (2012) developed a drilling parameter optimization system based on ANN to inform rig-site personnel in real-time to maximize the run length for all bits and downhole tools at maximum achievable ROP. The system uses drilling parameters obtained from offset log data. Additionally, prediction changes were imported to the ANN algorithm to improve the accuracy of future predictions. It is stated improved tool life, less trips, and conserving bit dull condition as the benefits of the system.

Irawan et al. (2012) used BYM to study the effects of several parameters on drilling operations. The most important parameters indicated as depth, pore pressure, equivalent circulating density, bit weight, rotary speed, bit tooth wear, and jet impact force. At the end of the study, the multiple regression analysis results were used to estimate the best WOB that gives optimum drilling operation. The ROP for each

specific depth intervals were predicted by using a field data, and optimized WOB was calculated for several data selections. In addition, a drilling simulator called Drill-Sim 500 was used to prove the results performed on actual field data.

Monazami et al. (2012) presented an application of ANN on predicting ROP by using several drilling data taken from an actual oil field. It is stated that ANN is a useful method to correlate complex parameters. Moreover, it is mentioned that this method can be used not only during well planning but also during drilling. To estimate the correlation coefficients a three layer feed-forward network has been developed. The results shows that ANN method can predict ROP with higher accuracy than conventional methods like BYM.

Jacinto et al. (2013) investigated several techniques on prediction and optimization of ROP. The techniques were a Bayesian approach and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS). They used their systems on a pre-salt region in Brazilian offshore. They also indicated that native Bayes is not a good approach on ROP prediction because of the classification of the input values, low data quality and division of the input pattern's nodes. On the other hand, DENFIS was capable of predicting ROP with low error. However, it is stated that the constraints of the system should be held by an expert due to extreme complexity.

Jamshidi and Mostafavi (2013) developed two models by utilizing ANN. The first model is selecting correct bit by optimizing ROP and applying particular drilling parameters. The second model is using appropriate drilling parameters acquired from optimization stage to select the bit that maximizes ROP. Additionally, GA is applied to the models. The methods let current condition of a drilling system optimize the drilling efficiency by reducing the probability of early wear of the bit.

Bilgin et al. (2014) performed in situ field tests to investigate the factors affecting drill bit wear mechanism and to obtain ideas about quality maintenance. The tests were performed under specific conditions in terms of rock formation characteristics and drill bit properties. They recorded bit wear radius until a threshold value was reached, then the bit were repaired and reused. The lifetime of the bits were expressed as achieving 29 mm of radius, which is the maximum allowable size. The results of the study is plotted in terms of drill bit diameter and drilling distance, which is shown in Figure 2.6.

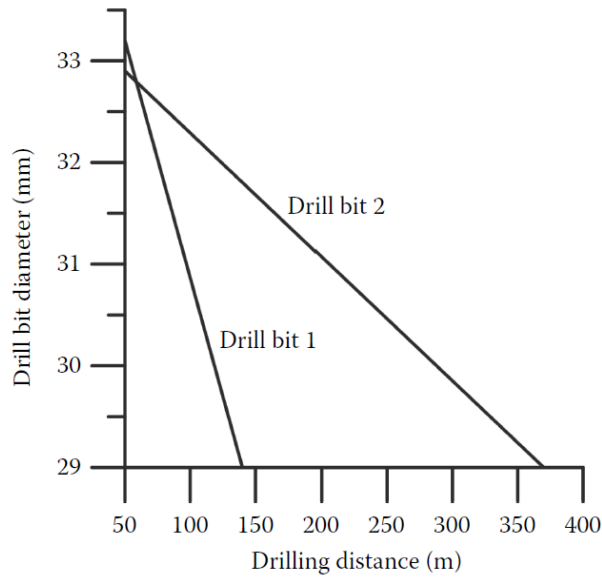


Figure 2.6 : Decrease in drill bit diameter via wear characteristics (Bilgin et al, 2014, p. 111).

Hou et al. (2014) conducted a study includes statistical analysis of a large number of actual drilling practices and detailed information about drilling bits. They researched the 3D drillability distribution of a drilling area and optimized drilling parameters using ANN based ROP model. Additionally, they applied a number of optimal bit selection techniques to the area. It is concluded that the ROP of a single well can be increased by 30% by applying these selection techniques in order to bring a solution to speed problems in deep wells.

Shishavan et al. (2014) combined both ROP and borehole pressure into a single controller system for a managed pressure drilling application. The system can control the drilling parameters such as pump flow rate, choke position, rotation rate, and WOB simultaneously. It is stated that stabilization of borehole pressure and maximization of ROP can be handled with automated operations. It is also reported that combining ROP and borehole control in a particular system can minimize risk, drilling costs, and operator workload. Additionally, consistency and predictability of gas influx reaction in shorter time is also one of the beneficial results of ROP and borehole combined system.

Miyora et al. (2015) presented a case study by mathematically modelling various drilling parameters such as formation strength, depth, formation compaction, pressure differential, bit diameter, WOB, bit rotation and bit hydraulics. The multiple linear regression was used as mathematical model. An actual field data was used for

developing the drilling model. Additionally, several software toolboxes were used in data analysis.

As an overview of the literature research based on ROP optimization, it can be said that the studies involve analytical, experimental, and statistical methods. Depending on the latest developments on computer technology, real-time analysis on the field has gained wide currency and has started to be used in automation processes to maximum ROP during drilling operations. However, despite all the methods and developments, there is no generalized solution for drilling parameters optimization. Besides, BYM is the most commonly used drilling model for ROP optimization, and provides basis to the recent studies.

In this study, a method that has not applied for the purpose of ROP optimization – SVR is used. There are several fields of application of SVM in the petroleum industry such as stuck pipe prediction, well log analysis for the estimation of saturation, lithology, porosity and permeability, PVT and reservoir property estimation, multiphase flow pattern recognition, seismic anomaly determination, history matching, reservoir simulation, etc. SVR and SVM has gained importance in the studies of petroleum industry in recent years. The following figure shows the number of papers found in OnePetro database that SVM and SVR were used in.

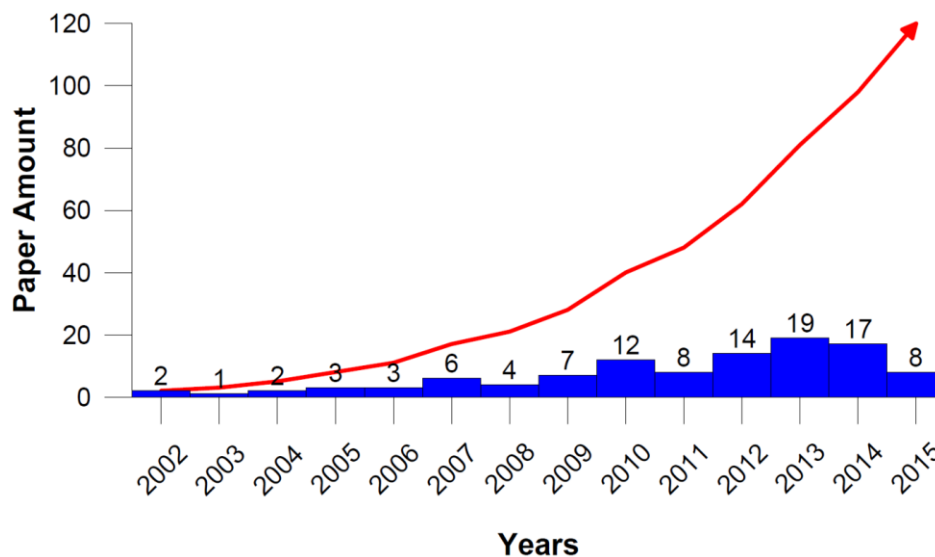


Figure 2.7 : Number of papers including SVM found in OnePetro database.

It can be seen from Figure 2.7 that there is a sharp increase in the number of papers after 2009. Additionally, the trend in paper amount shows exponential increase. This means that the studies on SVM for petroleum industry will be increasing.

3. STATEMENT OF THE PROBLEM AND SCOPE OF THE THESIS

It is a fact that there exists a tight relationship between maximizing rate of penetration (ROP) and reducing drilling cost. Therefore, accurate prediction of ROP both in planning phase and during drilling operation of a well is of great importance. There are many approaches and models introduced by numerous researchers, yet the one introduced by Bourgoyne Jr. and Young Jr. (1974) is one of the most commonly accepted and used by drilling industry. Bourgoyne and Young model (BYM) is also the one that is cited and examined by numerous researchers. The BYM estimates the ROP by considering eight common drilling parameters, which are depth, formation drillability, bit weight, rotary speed, tooth wear, mud hydraulics (Reynolds number function), equivalent circulating density, and pore gradient. The goal is to find eight correlation coefficients representing these eight drilling parameters by multiple regressing the offset data and to predict ROP for the desired conditions.

Theoretically speaking, estimating eight correlation coefficients from eight different parameters only eight input data is required. However, as stated by Bourgoyne Jr. and Young Jr. (1974) this mathematical expression is true if only the model represents the rotary drilling process with 100% accuracy (p. 374). Yet, it is a known fact that this situation never happens in real life. Additionally, it was also mentioned that negative values of coefficients could be faced with when a few input data were used for regression analysis (Bourgoyne Jr. and Young Jr., 1974, p. 374). Solving this kind of problem with lack of or unreliable data, a sensitivity study was performed by the researchers of the BYM study. To sum up, an overview of minimum number of input data points corresponding to parameter numbers must be available. As can be analysed mathematically, determining coefficients of eight parameter problem requires minimum of thirty data sets, and each data set comprises data of the eight parameters. As the parameter number decreases, the minimum required input data sets also decrease with a certain ratio.

Another issue is pointed out by BYM that the data should be taken from more than one well and combined to estimate the coefficients with maximum accuracy

(Bourgoyne Jr. and Young Jr., 1974, p. 375). Additionally, the data should be acquired in short depth intervals from one type of formation, especially shale. That is because most of the shale are stiff and non-brittle, which can resist fracturing and losing physical characteristics by bonding and oversteering (Yen and Chilingarian, 1976, p. 203). It is also stated that 2 to 5-ft is representative depth interval to keep the required data within reasonable limits (Bourgoyne Jr. and Young Jr., 1974, p. 375).

The situation above can be handled for the fields having large amount of shale sections like Offshore Louisiana Area as mentioned in the BYM study. However, the following situations can be faced within the most cases: (1) unavailability of adequate shale sections through one well to acquire enough drilling data, (2) not having sufficient number of wells in one field containing abundant shale sections, (3) drilling in a formation having extreme complexity in lithology, particularly in south eastern region of Turkey (Okay, 2008, p. 20).

Either one of these three cases results in restrictions on data acquisition or data insufficiency. When data is not adequate to regress and correlate the parameters, the reliability and confidence level of the regression analysis decreases. This situation causes unreasonable predictions for the target value or function, which is ROP.

On the other hand, there is a need for alternate approaches estimating coefficients of eight parameters if the input data set is insufficient (much less than thirty), or some of the data sets are not consistent. Of course, there are numerous models and methodologies for this purpose. As revealed in the previous studies on statistical analysis in various disciplines that Support Vector Machines (SVM), a machine learning model gaining more prominence particularly in the last decade, requires less input data to produce learning model and makes predictions more accurately rather than Artificial Neural Network (ANN), (Tolun, 2008). Moreover, Al-Baiyat and Heinze (2012) stated that the SVM is more useful and user-friendly than ANN, because the SVM requires fewer parameters to be optimized.

As detailed in the literature review chapter, the ANN brings more conservative prediction results in comparison to the BYM. Depending on the references given above, the SVM can be thought as a better predictor for ROP optimization problem. For these foundations, the applicability of SVM to optimize ROP is the main objective in this thesis study. By the best knowledge of this thesis author, the SVM is

investigated for the first time in the literature to predict ROP. The data provided by the study of Bourgoyne Jr. and Young Jr. (1974, p. 374) will be utilized for characterizing the effectiveness of SVM with respect to BYM.

4. MULTIPLE REGRESSION AND BOURGOYNE & YOUNG METHOD

In many cases of engineering applications, it is not possible to represent any dependent variable with only one independent variable. Engineering models are mostly the outcome of more than one subjects. Many variables can affect other variables by combination. These variables may also effect each other at the same time. For that reason, it is not achievable to perform simple regression analysis. According to Montgomery (2006), “A regression model that involves more than one regressor variable is called a multiple regression model.” (p. 63).

First part of this chapter mainly includes the mathematical background and different approaches of multiple regression analysis. In second part, the parameters and application of BYM is explained in details.

4.1 Theory of Multiple Regression Analysis

The general idea of multiple regression analysis is to estimate the linear or nonlinear relationship between dependent variable and independent variables, and to evaluate the effect of more than one independent variables in terms of response (Yan and Su, 2009, p. 41). Moreover, the computation of multiple regression analysis is more complex than simple regression, and the process may be problematic to interpret because of the relationships between independent variables (Freund et al, 2006, p. 73).

4.1.1 Introduction to multiple regression analysis

As stated before that multiple regression analysis involves more than one regressor variable. As an example, it can be supposed that Rate of Penetration (ROP) depends on only two variables such as rotary speed and WOB. A multiple regression model that can represent this relationship is expressed as,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (4.1)$$

Here, Y represents ROP, x_1 represents rotary speed, x_2 represents WOB, and ϵ is random error term. Equation (4.1) can be described as multiple linear regression model with two regressors. Additionally, this model is linear, because Equation (4.1) is a linear function of the unknown parameters β_0 , β_1 , and β_2 . β_0 is called as intercept, and β_1 and β_2 are called as partial regression coefficients, because β_1 is the measurement of the expected change in Y per unit change in x_1 when x_2 is constant (Montgomery and Runger, 2003, p. 411). In other words, Y is the dependent variable as x_1 and x_2 are independent or regressor variables. The generalized form of multiple linear regression model with k regressor variables is,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \quad (4.2)$$

Here, β_j ($j=0,1,\dots,k$) is called as regression coefficients, where the model expresses a hyperplane in the k -dimensional space of the regressor variables $\{x_j\}$ (Montgomery and Runger, 2003, p. 412).

4.1.2 Least squares estimation of the parameters

The least squares principle is used in multiple regression analysis to estimate the regression coefficients. Supposing there are $n > k$ observations expressed as,

$$(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) \mid i = 1, 2, \dots, n \mid n > k$$

Here x_{ij} denotes the i^{th} observation or input point of x_j . The data for multiple regression can be tabulated as shown in Table 4.1.

Table 4.1 : Data tabulation for multiple regression.

y	x_1	x_2	\dots	x_k
y_1	x_{11}	x_{12}	\dots	x_{1k}
y_2	x_{21}	x_{22}	\dots	x_{2k}
\vdots	\vdots	\vdots		\vdots
y_n	x_{n1}	x_{n2}	\dots	x_{nk}

If each observation is implemented in (4.2), the model forms as,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \quad (4.3)$$

and,

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i \mid i = 1, 2, \dots, n \quad (4.4)$$

The least squares principle for multiple regression forms as,

$$L = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (4.5)$$

It is desired to minimize L with respect to $\beta_0, \beta_1, \dots, \beta_k$ to get the minimum error (Montgomery and Runger, 2003, p. 414). The least squares estimates of $\beta_0, \beta_1, \dots, \beta_k$ must satisfy,

$$\left. \frac{\partial L}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) \quad (4.6a)$$

and,

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) x_{ij} \mid j = 1, 2, \dots, k \quad (4.6b)$$

By simplifying (4.6) the least squares normal equations are obtained as expressed in (4.7).

$$\begin{array}{ccccccc} n\hat{\beta}_0 & +\hat{\beta}_1 \sum_{i=1}^n x_{i1} & +\hat{\beta}_2 \sum_{i=1}^n x_{i2} & + \dots & +\hat{\beta}_k \sum_{i=1}^n x_{ik} & = \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} & +\hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 & +\hat{\beta}_2 \sum_{i=1}^n x_{i1}x_{i2} & + \dots & +\hat{\beta}_k \sum_{i=1}^n x_{i1}x_{ik} & = \sum_{i=1}^n x_{i1}y_i \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\beta}_0 \sum_{i=1}^n x_{ik} & +\hat{\beta}_1 \sum_{i=1}^n x_{ik}x_{i1} & +\hat{\beta}_2 \sum_{i=1}^n x_{ik}x_{i2} & + \dots & +\hat{\beta}_k \sum_{i=1}^n x_{ik}^2 & = \sum_{i=1}^n x_{ik}y_i \end{array} \quad (4.7)$$

It should be noted that there are $k + 1$ normal equations in the equation system, which include an unknown regression coefficient individually. The solution of (4.7) by any appropriate method will be resulted in regression coefficients.

4.1.3 Matrix approach

The multiple regression model can be expressed as a matrix notation as well. Former expression of a model including k regressor variable with n observations is shown in (4.3). The matrix notation of the same model is,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4.8)$$

where,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_n \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

It should be noted that $\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ are vectors including $(n \times 1)$ of observations, $(n \times p)$ of regressor variables, $(p \times 1)$ of regression coefficients, and $(n \times 1)$ of random errors, respectfully.

According to least squares principle, the goal is finding the vector of least squares estimators $(\hat{\boldsymbol{\beta}})$ that minimizes (4.9).

$$L = \sum_{i=1}^n \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (4.9)$$

$\hat{\boldsymbol{\beta}}$ is the solution for $\boldsymbol{\beta}$ in the equations that can be defined as in (4.10).

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \mathbf{0} \quad (4.10)$$

The solution of the system is done via,

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (4.11)$$

Equation (4.11) is the matrix form of (4.7). To solve (4.11) the both sides of the equation is multiplied by the inverse of $\mathbf{X}'\mathbf{X}$. Therefore, the least squares estimate is,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (4.12)$$

The matrix form of the normal equations can be expressed as below:

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{i1}x_{ik} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{bmatrix}$$

In conclusion, the fitted regression model is expressed as,

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{ij} \mid i = 1, 2, \dots, n \quad (4.13)$$

And the matrix notation of the same model is,

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (4.14)$$

The difference between observation and fitted value is called as residual and expressed as,

$$e_i = y_i - \hat{y}_i \quad (4.15a)$$

or,

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \quad (4.15b)$$

4.2 Bourgoyne and Young Method

The drilling model introduced by Bourgoyne Jr. and Young Jr. (1974) is the most common method to predict ROP for given conditions during drilling operations in the industry. The model includes the effects of (1) formation strength, (2) formation depth, (3) formation compaction, (4) pressure differential across the hole bottom, (5) bit diameter and bit weight, (6) rotary speed, (7) bit wear, and (8) bit hydraulics on penetration rate.

The BYM is described as a functional relationship between ROP and several drilling parameters as shown in Figure 4.1.

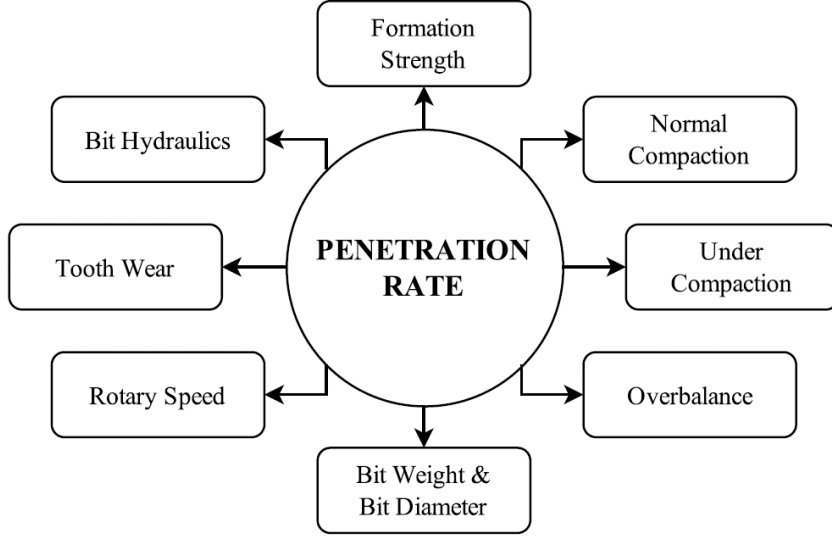


Figure 4.1 : Functional relationship of penetration rate.

The equation form of the functional relationship that shown in Figure 4.1 can be described as,

$$R = (f_1)(f_2)(f_3) \dots (f_n) \quad (4.16)$$

Each f term in (4.16) is expressed as a function of specific parameters. By considering multiple regression modeling principle, (4.16) can be presented as a form of (4.4).

$$R = \exp \left(a_1 + \sum_{j=2}^8 a_j x_j \right) \quad (4.17)$$

Here R is ROP, a_1 through a_8 are the regression coefficients that is determined by applying multiple regression analysis to offset data, and x_2 through x_8 are the functions related to several parameters.

4.2.1 Effect of formation strength

The effect of formation strength, or formation drillability, on ROP is represented directly by a_1 constant. It is stated by Bourgoyne Jr. and Young Jr. (1974) that formation strength is inversely proportional to the natural logarithm of square of the drillability constant introduced by Maurer (1962). Moreover, it also stands for non-modelled parameters such as mud type, and solid contents.

4.2.2 Effect of normal compaction

The exponential relationship between penetration rate and depth of a normally compacted formation is introduced by microbit studies by Murray and Cunningham (1955). The exponentially natural trend is also pointed out by field data in the study of Combs (1968). The relationship between relative penetration rate and vertical well depth is plotted on Figure 4.2.

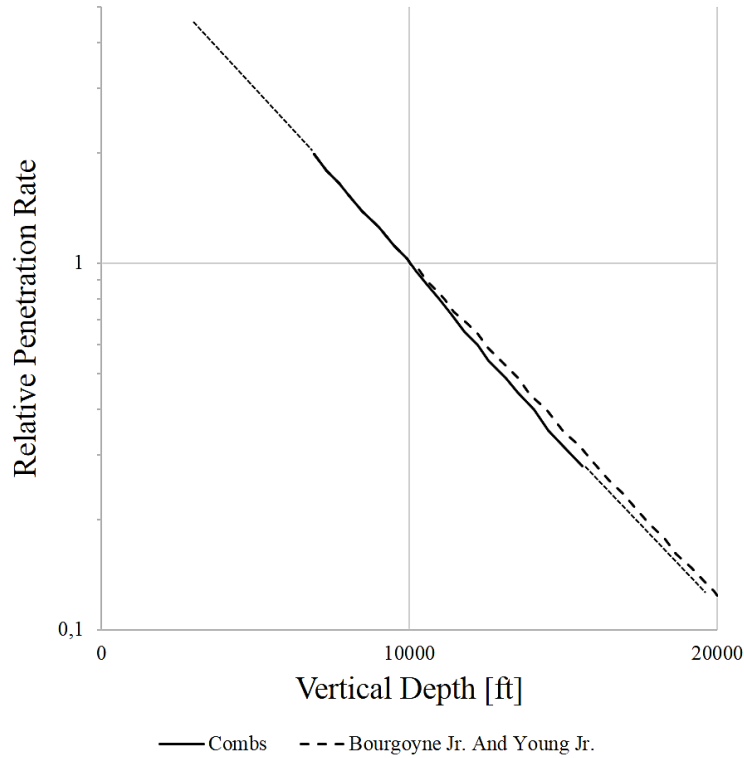


Figure 4.2 : Effect of normal compaction on ROP, adapted from (Bourgoyne Jr. and Young Jr., 1974, p. 372).

The effect of normal compaction is defined as,

$$x_2 = 10000 - D \quad (4.18)$$

Here D is well depth [ft]. It should be noted in (4.18) the term $e^{a_2 x_2}$ is normalized to be 1 at 10000 ft for a normally compacted formation.

4.2.3 Effect of under compaction

Under compaction model is the effect of abnormally pressured formations. It is assumed that there is an exponential increment in ROP with increasing pore pressure gradient. It is mentioned that the exponential nature of under compaction is based on

the compaction theory (Bourgoyne Jr. and Young Jr., 1974, p. 372). The effect of under compaction is defined as,

$$x_3 = D^{0.69}(g_p - 9) \quad (4.19)$$

Here g_p is the pore pressure gradient of the formation [lb/gal]. It should be noted that the under compaction exponent is normalized to the formations having 9 lb/gal pore pressure gradient. In other words, the term $e^{a_2x_2+a_3x_3}$ is going to be 1 for either a depth of 10000 ft or a pore pressure gradient of 9 lb/gal.

4.2.4 Effect of overbalance

Effect of overbalance can be described as the effect of pressure differential across the bottom hole on the ROP. According to the several studies performed by using field data and laboratory analyses, it is understood that there is an exponential relationship between ROP and excessive bottom hole pressure up to about 1000 psi (Vidrine and Benit, 1968; Combs, 1968; Cunningham and Eenink, 1958; Garnier and van Lingen, 1959). An example plot of the exponential relationship between pressure difference and ROP is shown in Figure 4.3.

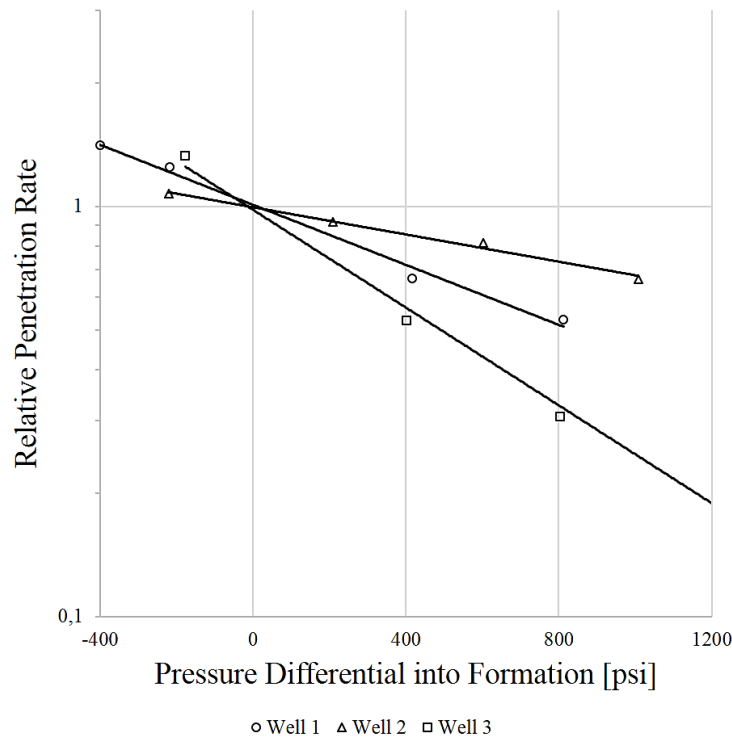


Figure 4.3 : Effect of differential bottom-hole pressure on ROP, adapted from (Bourgoyne Jr. and Young Jr., 1974, p. 372).

The overbalance term of the BYM is expressed as,

$$x_4 = D(g_p - \rho_c) \quad (4.20)$$

Here ρ_c is the equivalent circulating mud density at the hole bottom [lb/gal]. It can be seen from (4.20) that the value of $e^{a_4 x_4}$ is 1 when there is no overbalance.

4.2.5 Effect of bit weight and bit diameter

The effect of bit weight and diameter has been studied by several researchers either in laboratory or field. According to Bourgoyne et al. (1991), the general approach is to plot the relationship between ROP and bit weight gathered experimentally while holding all other drilling parameters as constants. The typical relationship between ROP and bit weight is shown in Figure 4.4.

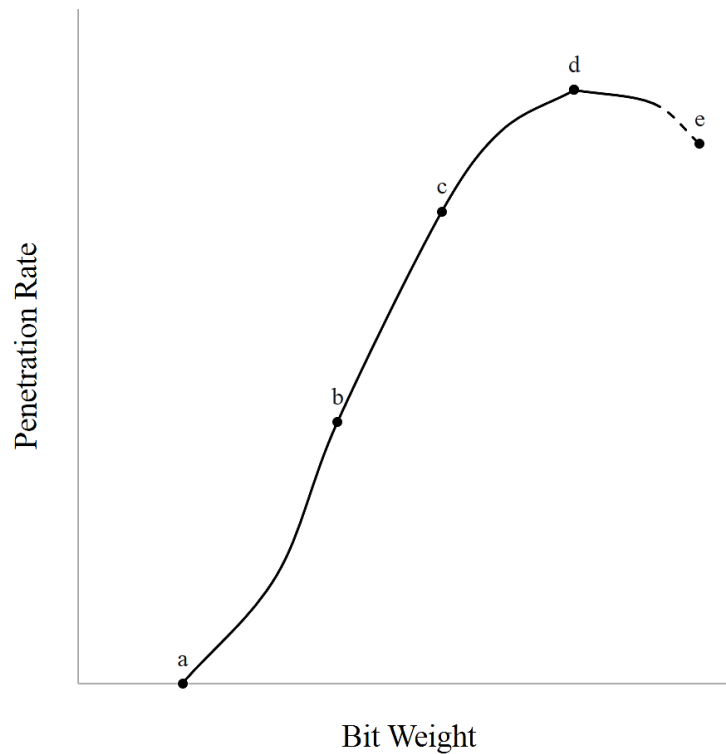


Figure 4.4 : Typical response of penetration rate to increasing bit weight, adapted from (Bourgoyne et al, 1991, p. 226).

In Figure 4.4, point (a) means there is no penetration rate until the threshold bit weight is reached; section (ab) represents a rapid increase in ROP with increasing moderate values of bit weight; section (bc) points out a linear trend between ROP and moderate bit weight; section (cd) shows more increase in bit weight produces a

slight increment on ROP; and section (de) declares excessive bit weight causes a decrease in ROP, which is called as bit floundering (Bourgoyne et al, 1991, p. 226).

It is also stated by Bourgoyne et al. (1991) high values of bit weight is related to low efficiency in bottom hole cleaning at higher rates of cutting transportation process (p. 226).

The effect of bit weight and bit diameter on ROP is formulated as,

$$x_5 = \ln \left(\frac{\frac{W_b}{d_b} - \left(\frac{W_b}{d_b}\right)_t}{4 - \left(\frac{W_b}{d_b}\right)_t} \right) \quad (4.21)$$

Here, W_b/d_b is the bit weight per inch of bit diameter [1000 lb/in], and $(W_b/d_b)_t$ is the threshold bit weight at which the bit begins to drill [1000 lb/in]. It is seen from (4.21) that the term $e^{a_5 x_5}$ is normalized to equal 1 for 4000 lb/in of bit diameter.

The threshold bit weight is determined via drill-off tests at very low bit weight values. Drill-off tests, which provide the desired operational conditions to optimize ROP and bit life, aim to yield the highest ROP available with an applicable bit wear rate (Lyons, 2010, p. 315). The primary objective in drill-off tests is not maximizing ROP because of the probability of bit life decrease. A drill-off test, which involves applying a large amount of weight to the bit, then locking the brake and monitoring the reducing in bit weight with constant rotary speed, determines the response in ROP by changing bit weight in a short depth intervals by the application of Hook's Law (Bourgoyne et al., 1991, p. 227).

Additionally, it is mentioned by Bourgoyne et al. (1991) that the threshold bit weight usually has negligible small values in the areas having relatively soft formations (p. 234). In this study, the threshold bit weight is assumed as zero, as in the original BYM study.

4.2.6 Effect of rotary speed

Similar to the effect of bit weight, the effect of rotary speed is directly proportional to ROP. The effect of rotary speed can be figured by plotting ROP vs. rotary speed while all other drilling parameters assumed as constants. An example plot is seen on Figure 4.5.

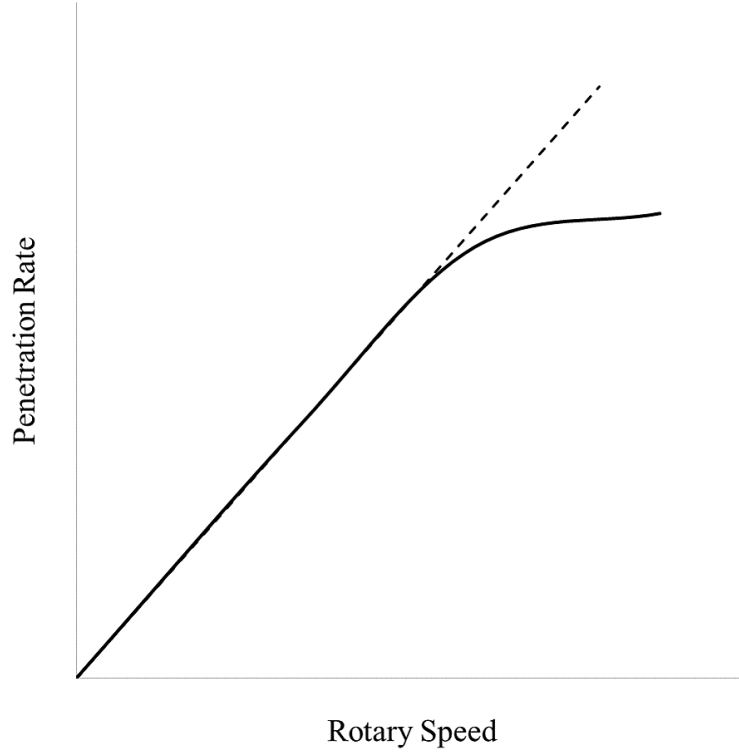


Figure 4.5 : Typical response of penetration rate to increasing rotary speed, adapted from (Bourgoyne et al, 1991, p. 226).

As seen in Figure 4.5, the ROP generally has a linear relationship with increasing rotary speed at low values. At higher rotary speed values, the ROP tend to decrease after a point. This behaviour is related to less efficient bottomhole cleaning (Bourgoyne et al, 1991, p. 226).

The effect of rotary speed on ROP is expressed as,

$$x_6 = \ln\left(\frac{N}{100}\right) \quad (4.22)$$

Here N is rotary speed [rpm]. It is important that the term $e^{a_6 x_6}$ is normalized to be equal 1 at 100 rpm.

4.2.7 Effect of tooth wear

The effect of tooth wear was examined and modelled by researchers in the previous studies (Galle and Woods, 1963; Edwards, 1964). However, according to Bourgoyne Jr. and Young Jr. (1974), these models were complex to express and not suitable for performing multiple regression analysis (p. 372).

The effect of tooth wear on ROP is modelled as,

$$x_7 = -h \quad (4.22)$$

Here, h is the fractional tooth height worn away. In Figure 4.6, there is a comparison of tooth wear models mentioned before.

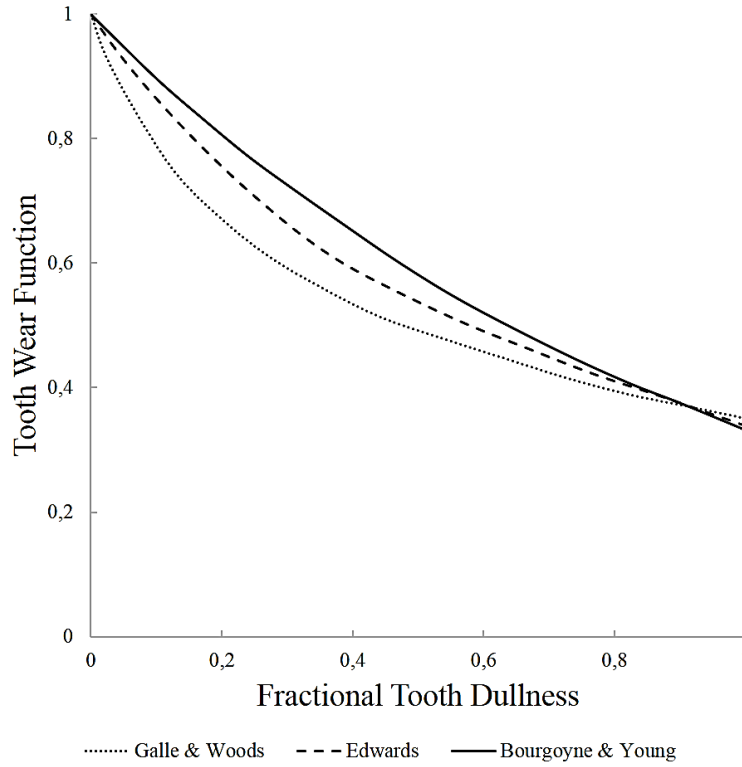


Figure 4.6 : Effect of tooth wear on penetration rate, adapted from (Bourgoyne Jr. and Young Jr., 1974, p. 373).

As seen on Figure 4.6, the effect of tooth wear according to BYM is the most convenient behaviour to model. It is also mentioned that, the value of tooth wear exponent mainly depends on the bit type and formation type (Bourgoyne Jr. and Young Jr., 1974, p. 372).

Basically, worn tooth height for a roller-cone bit is a measurement having a value between 0 to 8. As seen on Figure 4.7, 0 means new, and 8 means completely dull bit. Therefore, the fractional worn tooth height is the division of the average wear of the row of teeth to 8.

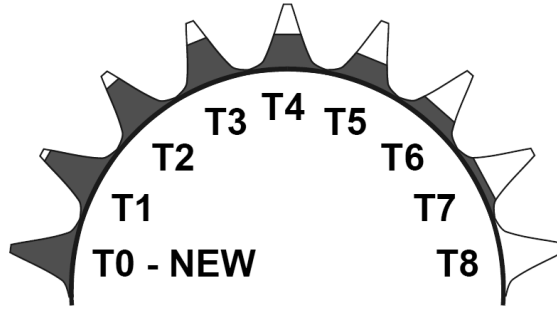


Figure 4.7 : Tooth wear chart for a roller-cone bit (Smith Bits, 2008).

In addition, the tooth wear model of BYM is valid for roller-cone bits. When Polycrystalline Diamond Compact (PDC) or Tungsten Carbide Core (TCC) bits are used, the change in ROP can be negligible (Bourgoyne Jr. and Young Jr., 1974, p. 372). Thus, in these kind of situations the regression coefficient of tooth wear function (a_7) is assumed to be zero.

4.2.8 Effect of bit hydraulics

The effect of bit hydraulics was first based on the microbit experiments performed by Eckel (1967), which includes a Reynolds number group. The old model is expressed as,

$$x_{8,old} = \frac{\rho q}{350\mu d_n} \quad (4.23)$$

Here, ρ is mud density [lb/gal], q is flow rate [gal/min], μ is the apparent viscosity at 10000 sec^{-1} [cp], and d_n is bit nozzle diameter [in]. The effect of the old model on ROP is seen on Figure 4.8.

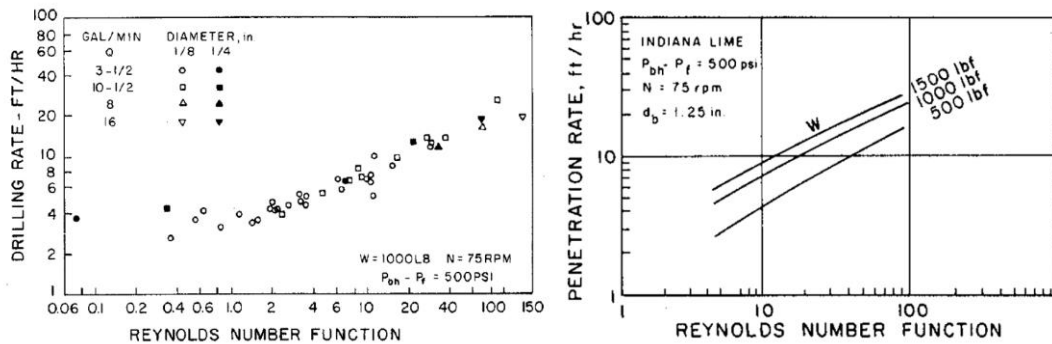


Figure 4.8 : ROP as a function of bit Reynolds number, (Bourgoyne et al, 1991, pp. 230, 231).

Figure 4.8 is a summary of the study of Eckel (1967). It is seen that the ROP and Reynolds number function is directly proportional to each other. If the bit weight is increased, the trend is simply shifted upwards. However, the behaviour at the bit flounder point is not involved in Eckel's study. Thus, as mentioned by Bourgoyne et al. (1991), Eckel's model is not applicable to be used in field practice (p. 232).

It is stated that hydraulic horsepower and jet impact force is the most commonly used models to develop the correlations between bit hydraulics and the ROP (Bourgoyne et al, 1991; Wright et al, 2003). In addition, it is indicated by Bourgoyne et al. (1991) that the jet Reynolds number group, hydraulic horsepower, and jet impact force give similar results for correlating the effect of bit hydraulics on the ROP (p. 232). The basis of this outcome is the full-scale acquired data from the study of Tibbitts (1991), which was performed under simulated borehole conditions in drilling laboratory experiments. The summary of this study is seen on Figure 4.9.

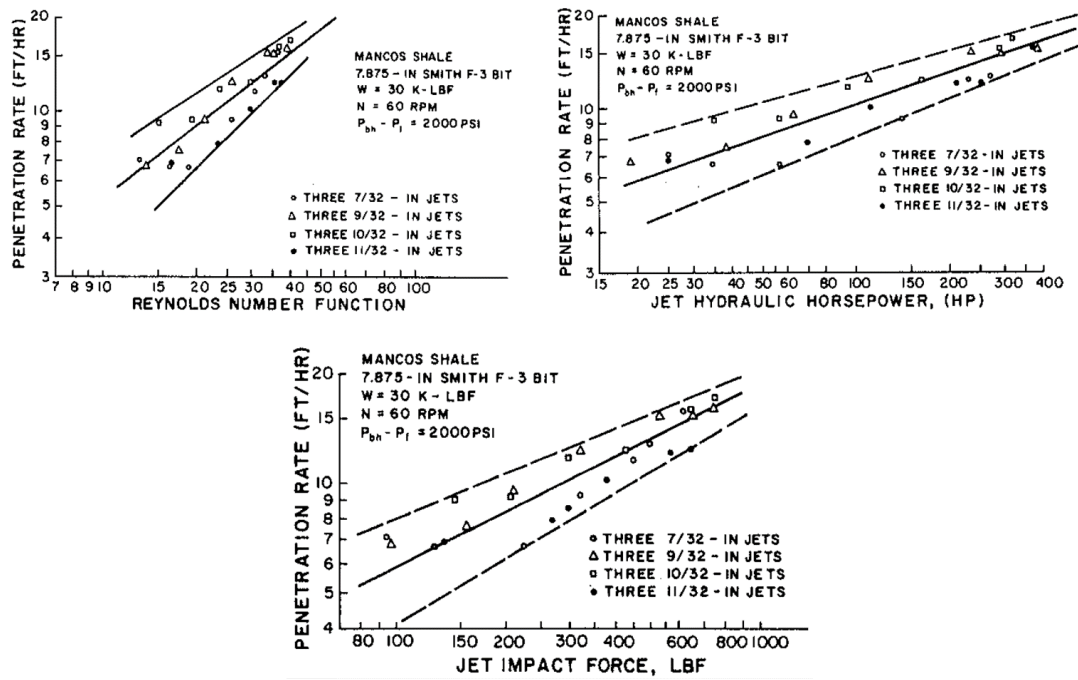


Figure 4.9 : The effect of different hydraulic models on ROP (Bourgoyne et al, 1991, p. 231).

In Figure 4.9, it can be seen the effect of different hydraulic models on the ROP in a field, under simulated borehole conditions. Hence, the effect of bit hydraulic on the ROP is formulated as,

$$x_8 = \ln\left(\frac{F_j}{1000}\right) \quad (4.24)$$

Here F_j is the hydraulic impact force beneath the bit [lbf]. The effect of jet impact force is normalized to be equal to 1 at 1000 lbf.

It should be noted that, as seen on Figure 4.9, the behaviour of ROP with changing variables of Reynolds number function and jet impact force is similar. The data used in this thesis study involves Reynolds number function instead of jet impact force. Thus, the Reynolds number function is used for calculating the exponent of the effect of bit hydraulics.

5. MACHINE LEARNING AND SUPPORT VECTOR REGRESSION

Nowadays, computers have the capability of both making decisions about events and learning relationships between occurrences. The problems, which are impossible to formulate and solve mathematically, can be solved by computers via intuitive methods. The concept that equips and develops computers with these properties is called as artificial intelligence (Shi, 2011).

The main point of artificial intelligence studies is to let computers solve the problem when there is no such an algorithm that solves the current problem. In order to manage this situation, all the data and information about the statement of the problem must be given to the computer previously. The accuracy and reliability of results depend on the number, righteousness and validity of the input data (Negnevitsky, 2002).

Data acquisition can be done via several approaches. For example, questionnaires, literature surveys, meetings with experts, interviews, working with specialists, or gathering information from previous samples related to current occurrence (learning from samples). The process of analyzing and extracting information from data is called as data mining (Witten and Frank, 2005).

There are several artificial intelligence techniques that used to gather information. Expert Systems are the technology that develops computer software, which solve a problem, as a specialist will do (Malhotra, 2001). Expert Systems have four basic elements, which are procuring of data, database, estimation mechanism, and a user interface. The flowchart of the operation procedure of Expert Systems is shown in Figure 5.1.

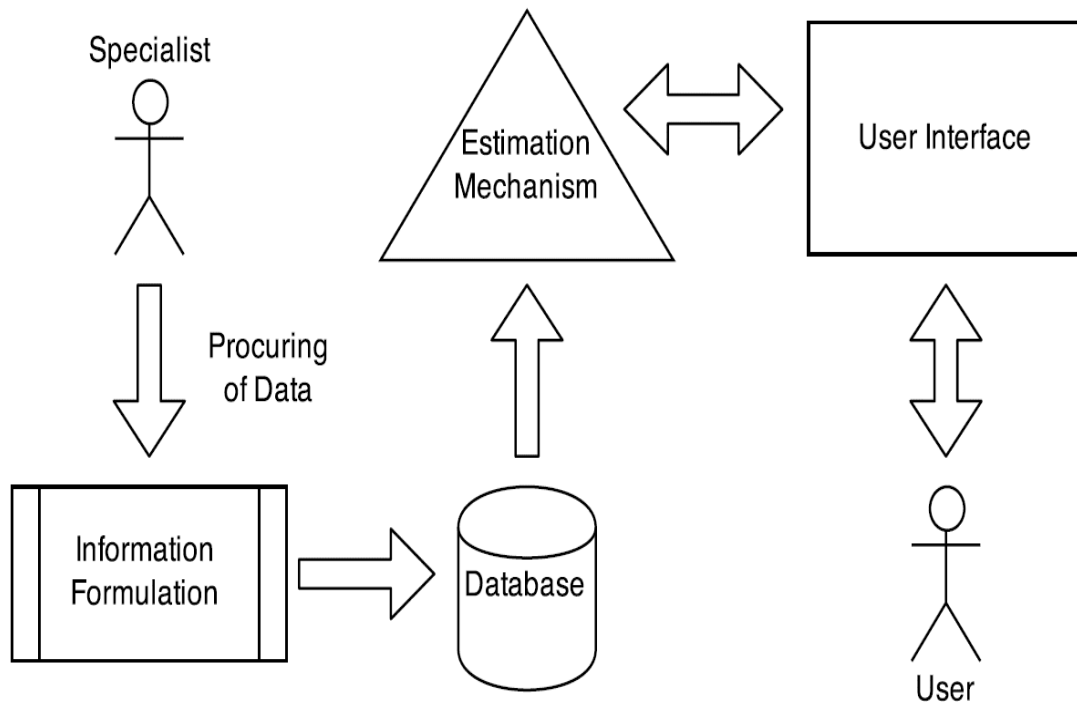


Figure 5.1 : Expert Systems flowchart.

Genetic Algorithms (GA) is another commonly used artificial intelligence technique to solve complex optimization problems. In GA, first, random searches are performed to solve a problem. Then, these searches are matched together to generate high performance solutions. This procedure continues until there is no better result generated.

There are several notions and parameters in GA needs to be clarified. “Chromosomes” are each individual solutions of the problem that is needed to be solved by GA. One problem can has N solutions. GA is desired to find the best of these N solutions. Each element of a chromosome represents a unique property of the solution, which is called “gene”. “Search Space” is the first solution set to be used for searching the best solution. “Crossover” is generating new solution by pairing chromosomes. “Mutation” is changing a gene to make searching new solution easier and to change the direction of the search (Sivanandam and Deepa, 2008). Figure 5.2 shows the flowchart of GA.

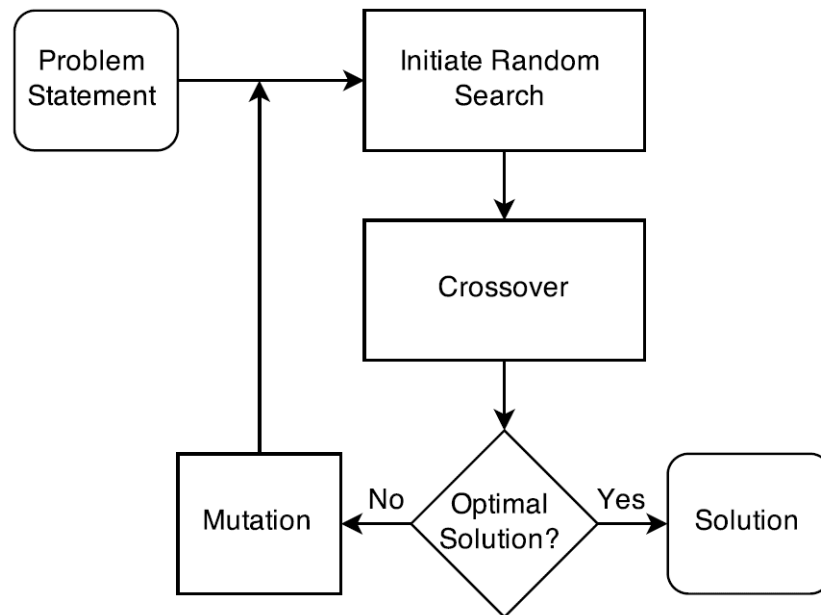


Figure 5.2 : Genetic Algorithms flowchart.

Another method for artificial intelligence applications is the “Fuzzy Logic”. Most of the events occur in uncertain conditions. Unexpected situations may happen and affect decision making. Fuzzy Logic is the technology developed for let the computer decide in these circumstances. As a basis, Fuzzy Logic consists of three steps. First, fuzzifying, which is the process of determination of fuzzy proposition variables, and rule bases for the problem to be solved. Second, inference, which is estimation of solution area by using the rule bases of determined fuzzy proposition variables. Finally, defuzzifying, which is gathering only one value from estimated solution area (Sivanandam et al, 2007). The flowchart of a Fuzzy Logic algorithm is shown in Figure 5.3.

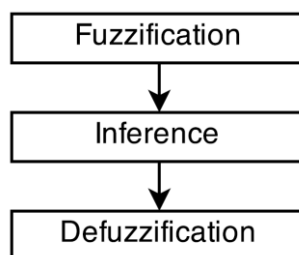


Figure 5.3 : Fuzzy Logic flowchart.

The most common and recent method among the artificial intelligence techniques is “Machine Learning”. Machine Learning (ML) is making computer learn from data. In ML, computers learn the information and experiences about instances by data

mining, decide about future occurrences and generate solutions for similar problems. Thus, ML applications can be met in many fields during daily life. ML related fields is shown in Figure 5.4

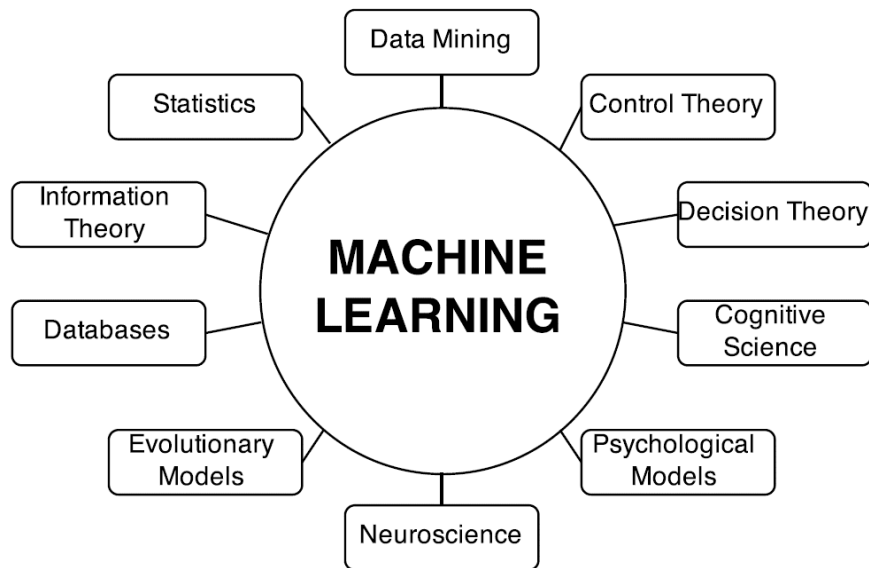


Figure 5.4 : Machine Learning related fields.

In ML, there is a binary classification process called as “task” to assign the objects to different classes. It is desired to establish a connection between features and classes, which is called as “model” by analysing “training sets” (Flach, 2012, p. 11). Figure 5.5 shows the relation between the elements of ML.

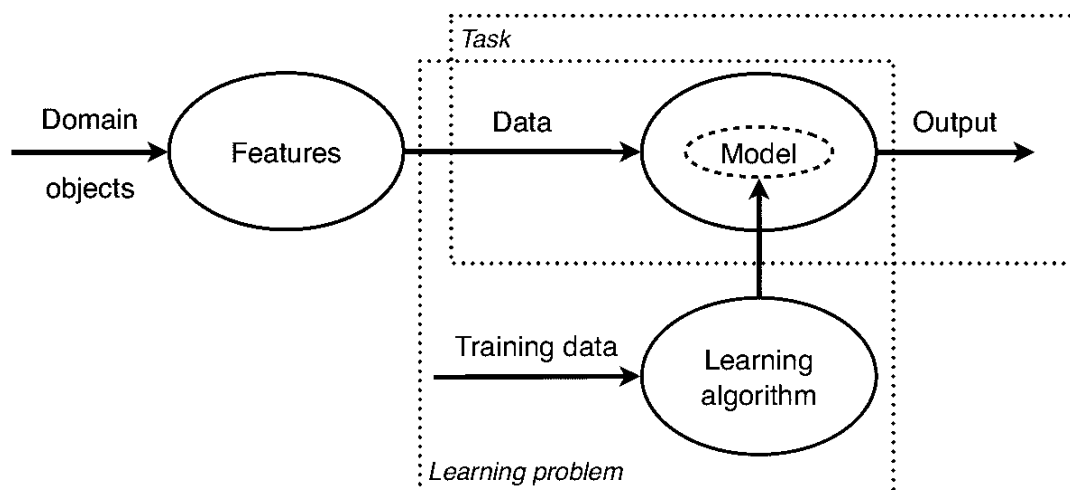


Figure 5.5 : An overview of how ML works (Flach, 2012, p. 11).

In Figure 5.5, it is seen that in order to establish a task, a proper model mapped from data is required. The mapping process from a training dataset forms the learning

problem. In short, according to Flach (2012), “Tasks are addressed by models, whereas learning problems are solved by learning algorithms that produce models. Machine learning is concerned with using the right features to build the right models that achieve the right tasks.” (p. 12).

There are mainly two settings for ML. First setting is Supervised Learning, which is called for the task to build an input-output relationship from a labelled training set (Zhang, 2001, p. 103). And the second is Unsupervised Learning, which is the learning process of the algorithm from unlabelled data (Ghahramani, 2004, p .3). The ML settings is summarized in Table 5.1.

Table 5.1 : Summary of ML settings, adapted from (Flach, 2012, p. 18).

	Predictive Model	Descriptive Model
Supervised Learning	classification, regression	Subgroup discovery
Unsupervised Learning	predictive clustering	Descriptive clustering, Association rule discovery

ML methods divided into three subgroups in terms of modelling. First model is geometric model, which considers instances as in geometrical concepts such as coordinate systems, lines, planes, and distances (Mulmuley, 1994, p .26). Figure 5.6 shows an example of geometric distribution of instances on a 2-dimensional space.

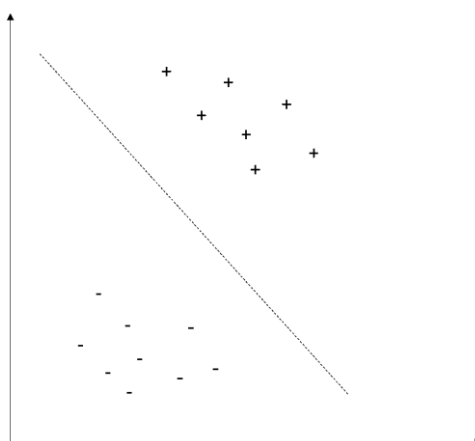


Figure 5.6 : An example of linear classification in two dimensions.

The second model is probabilistic model, which behaves as a Bayesian classifier in nature (Flach, 2012, p. 25). The main idea of this model is to transform two basic approaches, posterior probabilities and likelihoods, by using Bayes’ rule. The Bayes’ rule is expressed as,

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (5.1)$$

Here P(Y) is the prior probability and P(X) is the probability of the data.

The third type of model is logical model, which can be allowed to be translated into rules that people can understand and easily organised as a tree-structured decision mechanism (Flach, 2021, p. 32). An example of a decision tree is shown in Figure 5.7.

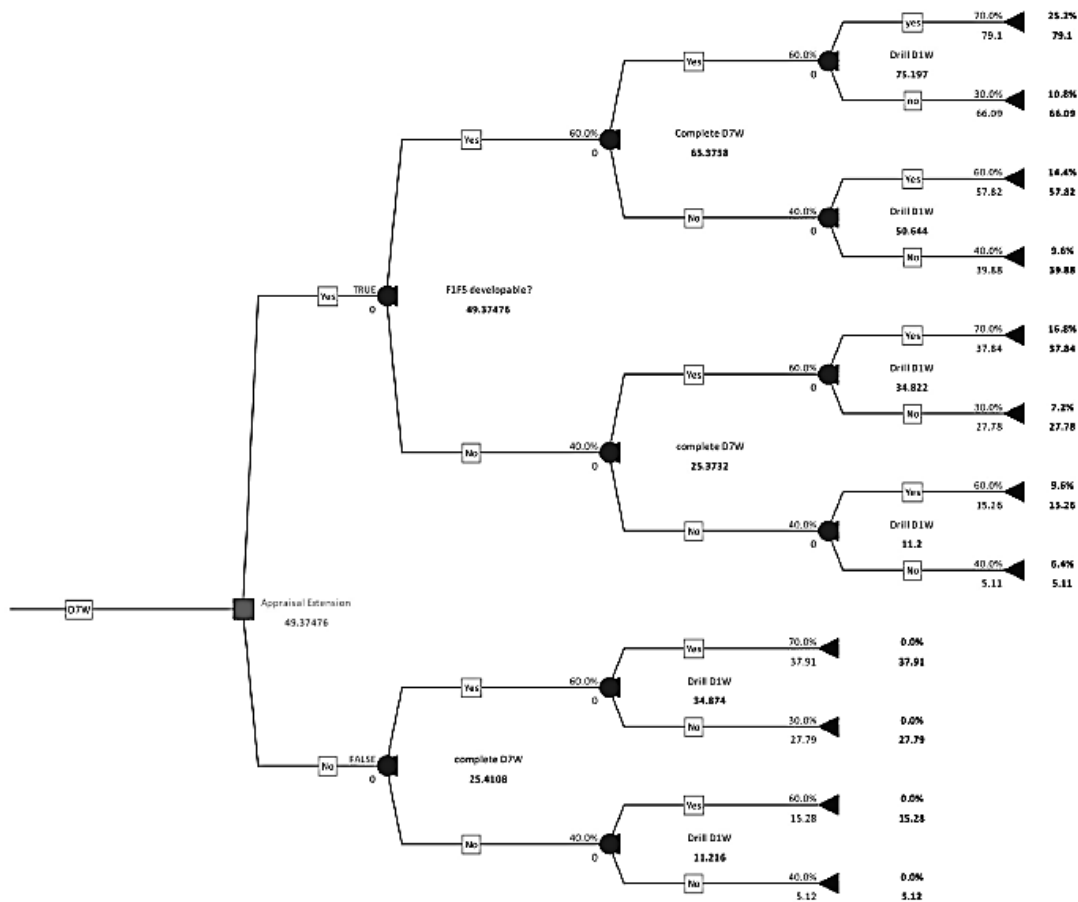


Figure 5.7 : An example of decision tree (Cyril et al, 2014, p. 5).

To sum up and map the models of ML methods, the diagram shown in Figure 5.8 was introduced by Flach (2012).

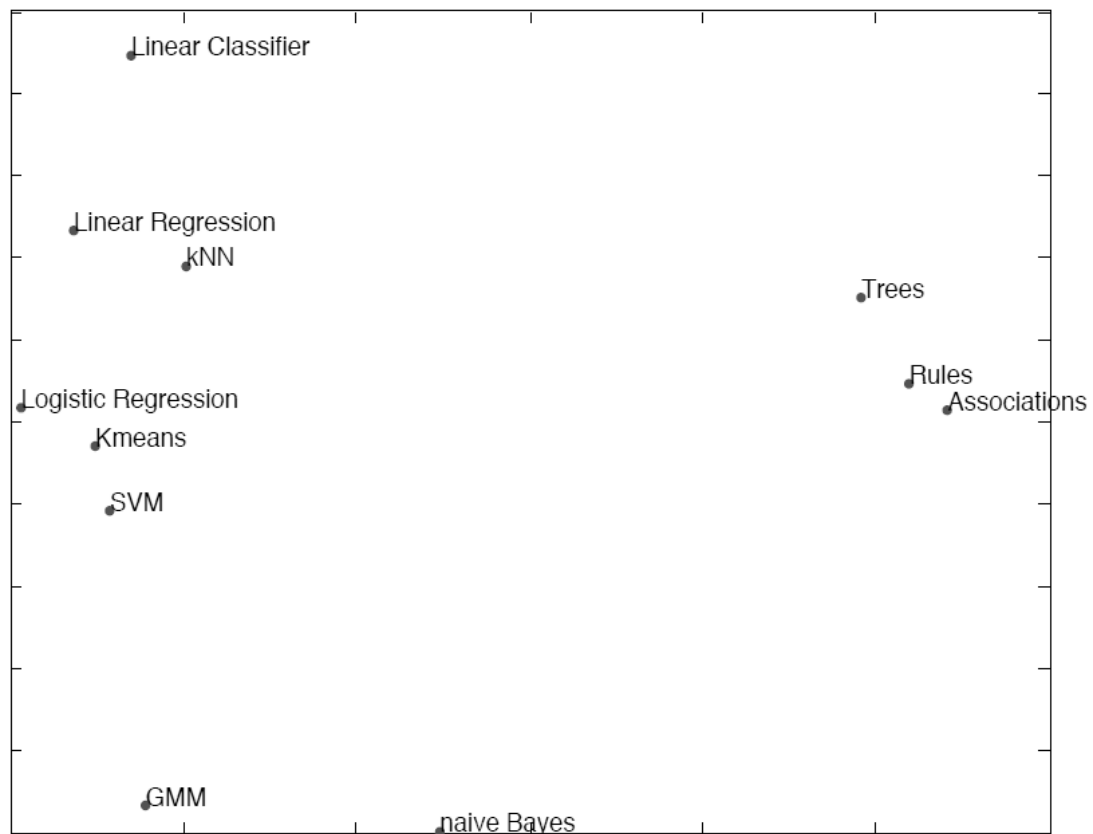


Figure 5.8 : Mapping of ML methods (Flach, 2012, p. 37).

In Figure 5.8, the models that have same characteristics are plotted closer. Geometric models are on the top left, probabilistic models are on the bottom left, and logical models are on the right in the diagram.

Among the geometric, probabilistic and logical models of ML, there is another abstract dimension that supports all of the ML models. This dimension can be defined as the perception between grouping and grading models. The main divergence between these models is the approach to handle the instance space. Grouping models divide instance space into segments, while grading models form a global model on the instance space (Flach, 2012, p. 36).

To combine the settings and models of the ML methods, the listing in Figure 5.9 was represented by Flach (2012).

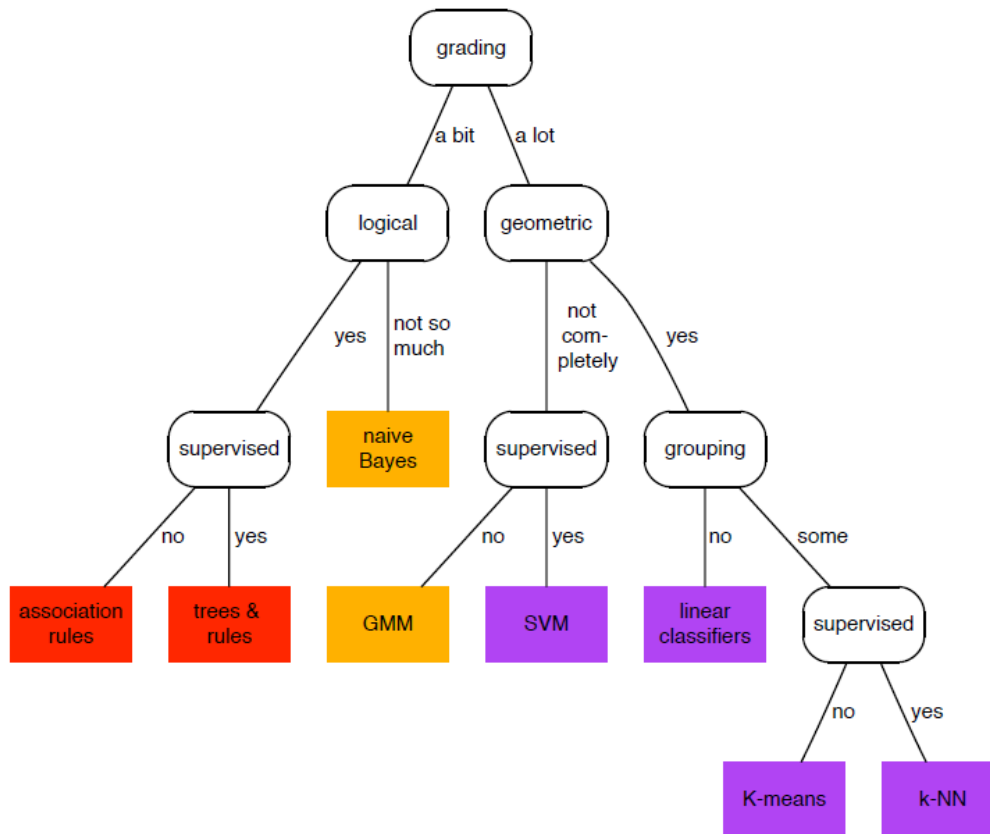


Figure 5.9 : ML taxonomy (Flach, 2012, p. 38).

In Figure 3.8, there is a taxonomy that describes ML methods in terms of expansion to the ability of grading or grouping, model of logical, geometric or combination, and setting of supervised or unsupervised.

In this thesis study, a supervised and geometric method of ML, Support Vector Machines (SVM), is used. The basic mathematics and the fundamental idea of SVM is explained in the following topics.

5.1 Statistical Learning Theory

Statistical Learning Theory was introduced by Vapnik (1998). The purpose of modelling is choosing a model from the hypothesis space that is the closest to the main function in the target space with a measurable error (Gunn, 1998, p. 2).

There are two kind of errors occurring from two cases. First error is the approximation error, which is the measurement of the ability of functions to approach the target (Bousquet et al, 2004, p. 181). The second error is called as estimation error, which is the error of the learning process that occurred via

prediction rules based on a set of random examples (Kulkarni and Harman, 2011, p. 14). These errors together create generalization error. The schematic presentation of the errors is shown in Figure 5.10.

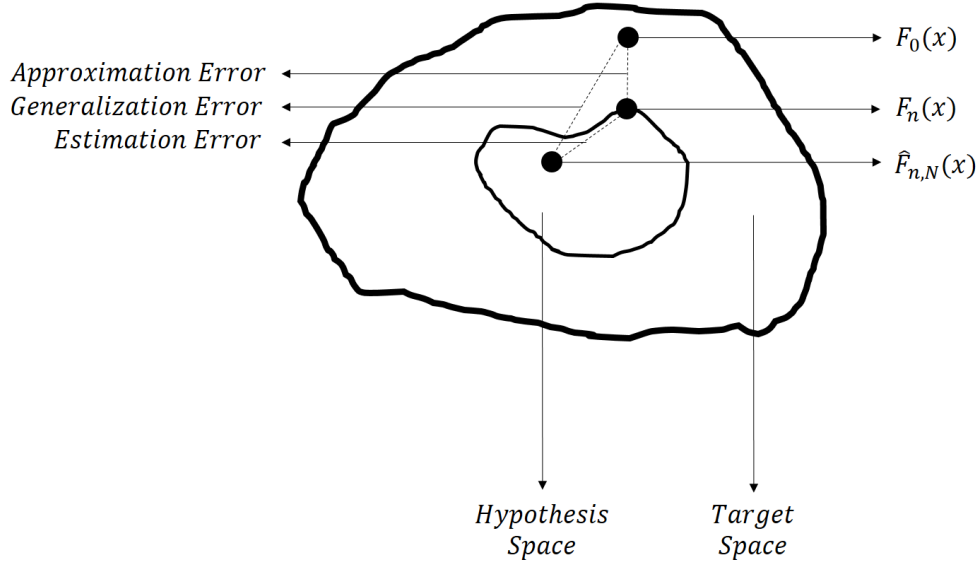


Figure 5.10 : Modelling errors.

5.1.1 VC dimension

Vapnik-Chervonenkis (VC) dimension is a scalar, which is a measure of the capacity of a function set. In other words, VC dimension is a scalar value, which can be calculated for any function set applied to a learning machine (Vapnik, 1999, p. 69). It is defined by Gunn (1998) that “The VC dimension of a set of functions is p if and only if there exists a set of points $\{x^i\}_{i=1}^p$ such that these points can be separated in all 2^p possible configurations, and that no set $\{x^i\}_{i=1}^q$ exists where $q > p$ satisfying this property.” (p.4). An example of VC dimension is shown in Figure 5.11.

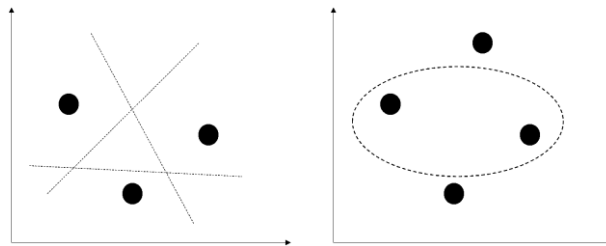


Figure 5.11 : VC dimension illustration.

In Figure 5.11, the ability of the set of linear indicator functions to shatter three points in a plane is illustrated. It should be noticed that these linear functions cannot distribute four points. In this situation, the VC dimension is equal to the number of

free parameters whereas the set of linear indicator functions in n dimensional space has $n + 1$ VC dimension (Gunn, 1998, p. 4).

5.1.2 Structural risk minimisation

If a structure, which S_h is a hypothesis space of VC dimension (h_{VC}) is created as $S_1 \subset S_2 \subset \dots \subset S_\infty$, the Structural Risk Minimization (SRM) is about solving the following expression.

$$\min_{S_h} R_{emp}[f] + \sqrt{\frac{h_{VC} \ln\left(\frac{2l}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{l}} \quad (5.2)$$

where,

$$R_{emp}[f] = \frac{1}{l} \sum_{i=1}^l L(y^i, f(x^i)) \quad (5.3)$$

Here $L(y, f(x))$ is the loss between the input x with corresponding response y and the response $f(x)$ estimated by the learning machine (Vapnik, 1991, p. 832). Hence, $R_{emp}[f]$ term is the empirical risk function. The second term of the expression above is the confidence interval between the bounds of the actual risk and the empirical risk (Vapnik, 1991, p. 834).

It is stated by Gunn (1998) that multiple output problems have the capability of being reduced to a set of single output problems as an independent consideration, so that predicting a desired single output from multiple inputs is more convenient (p.4).

5.2 Support Vector Machines for Classification

Support Vector Machine (SVM) is a learning method, which bases on finding a hyperplane that separated the d -dimensional data in to its two classes perfectly, used for binary classification (Boswell, 2002, p. 1). The goal is to establish a classifier, which must be appropriate for unseen examples to generalize completely (Gunn, 1998, p. 5). An example of classifiers and hyperplanes is shown in Figure 5.12.

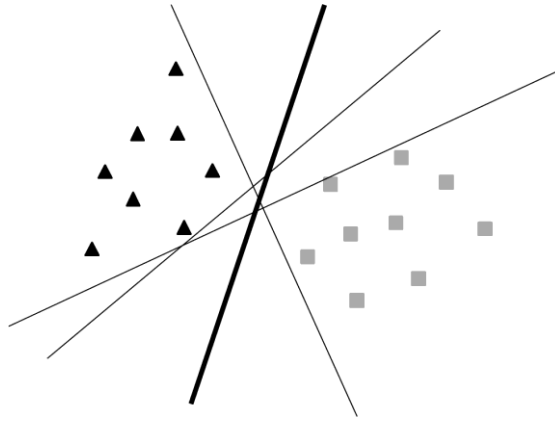


Figure 5.12 : Classifier examples.

As seen in Figure 5.12, there are several possibilities to separate the data with linear classifiers. However, there is only one classifier that maximizes the distance between itself and the nearest data point of each class, which is drawn in bold among other classifiers in Figure 5.12. The overview of classification concept is shown in Figure 5.12.

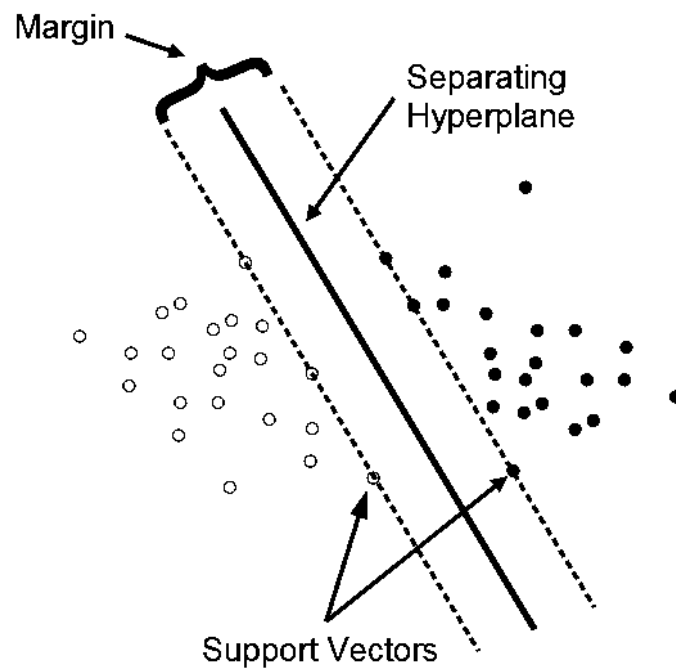


Figure 5.13 : Linear classification (Meyer, 2014, p. 2).

As shown in Figure 5.12, the nearest data points are named as “support vectors”, the distance between support vectors is called as “margin” and the classifier that maximizes the margin is called as “optimal separating hyperplane” (Gunn, 1998, p. 5).

5.2.1 The optimal separating hyperplane

The problem is to separate the set of training vectors associated with two separate classes with a hyperplane, as expressed in (5.4) and (5.5).

$$D = \{(x^1, y^1), \dots, (x^l, y^l)\} \mid x \in \mathbb{R}^n, y \in \{-1, 1\} \quad (5.4)$$

$$f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (5.5)$$

Here w is the weight vector, and b is bias. If b is considered as zero, the points of x that corresponds $\langle \mathbf{w}, \mathbf{x} \rangle = 0$, or in other expression $\mathbf{w}^T \mathbf{x} = 0$, lie perpendicular to \mathbf{w} and intersect origin. This situation represents a line in two dimensions, or a plane in three dimensions, which is generally named as “hyperplane”. The geometric meaning of b is the translation of how far the hyperplane away from the origin (Ben-Hur and Weston, 2009, p. 2).

$$\{\mathbf{x} : f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0\} \quad (5.6)$$

The hyperplane expressed in (5.6) divides the space into two sides. Here $f(\mathbf{x})$ represents the linear discriminant function. The sign of $f(\mathbf{x})$ stands for which side of the hyperplane a point is present.

If the error is assumed as zero, it can be said that the vectors are optimally separated by the hyperplane and the margin is maximum. Furthermore, it is stated by Vapnik (1995) that (5.6) includes repetitions and it is proper to apply an optimal canonical separating hyperplane (OCSH) without generality loss. Now, the constraint of the parameters can be described as,

$$\min_i |\langle \mathbf{w}, \mathbf{x}^i \rangle + b| = 1 \quad (5.7)$$

This strict constraint on the parameters helps in simplifying the formulation. According to Gunn (1998), “... the norm of the weight vector should be equal to the inverse of the distance, of the nearest point in the data set to the hyperplane.” (p. 6). An overview of OCSH is shown in Figure 5.14.

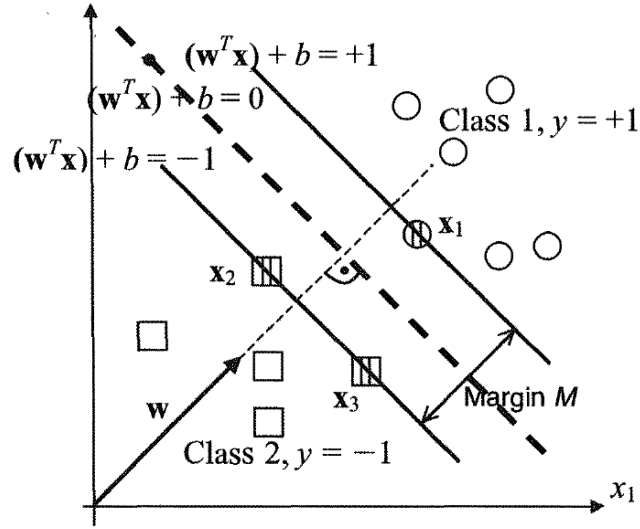


Figure 5.14 : Optimal canonical separating hyperplane (Kecman, 2001, p. 154).

In Figure 5.14, OCSH with maximum margin stays in the halfway of two classes. The closest points to the hyperplane are support vectors, which are x_1 from Class 1, and x_2 and x_3 from Class 2.

According to the geometric approach, an OCSH must recoup the following constraints:

$$y^i [\langle \mathbf{w}, \mathbf{x}^i \rangle + b] \geq 1 \mid i = 1, \dots, l \quad (5.8)$$

And the distance of a point (x) from the hyperplane is,

$$d(w, b; x) = \frac{|\langle \mathbf{w}, \mathbf{x}^i \rangle + b|}{\|\mathbf{w}\|} \quad (5.9)$$

Here l is the number of training data. As mentioned before, the optimal hyperplane is maximizes the margin (M) with respect to the constraints defined in (5.8). The margin is calculated explicitly in the following equations.

$$M = \min_{x^i: y^i = -1} d(w, b; x^i) + \min_{x^i: y^i = 1} d(w, b; x^i) \quad (5.10)$$

$$M = \min_{x^i: y^i = -1} \frac{|\langle \mathbf{w}, \mathbf{x}^i \rangle + b|}{\|\mathbf{w}\|} + \min_{x^i: y^i = 1} \frac{|\langle \mathbf{w}, \mathbf{x}^i \rangle + b|}{\|\mathbf{w}\|} \quad (5.11)$$

$$M = \frac{1}{\|\mathbf{w}\|} \left(\min_{x^i: y^i = -1} |\langle \mathbf{w}, \mathbf{x}^i \rangle + b| + \min_{x^i: y^i = 1} |\langle \mathbf{w}, \mathbf{x}^i \rangle + b| \right) \quad (5.12)$$

$$M = \frac{2}{\|\mathbf{w}\|} \quad (5.13)$$

Therefore, the optimal separating hyperplane minimizes,

$$\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (5.14)$$

As mentioned by Gunn (1998), (5.14) is independent of bias, because in (5.8) changing bias results in a movement in the normal direction to separating hyperplane itself (p. 7). Respectfully, the margin does not change. However, the hyperplane is not considered as optimal for this situation, because it is now nearer to one class than the other class.

To minimize (5.14), SRM is implemented by defining holding bounds as $\|\mathbf{w}\| < A$. Hence, the distance between the hyperplane and any of the data points cannot be nearer than $1/A$. Figure 5.15 demonstrates the constaining canonical hyperplanes.

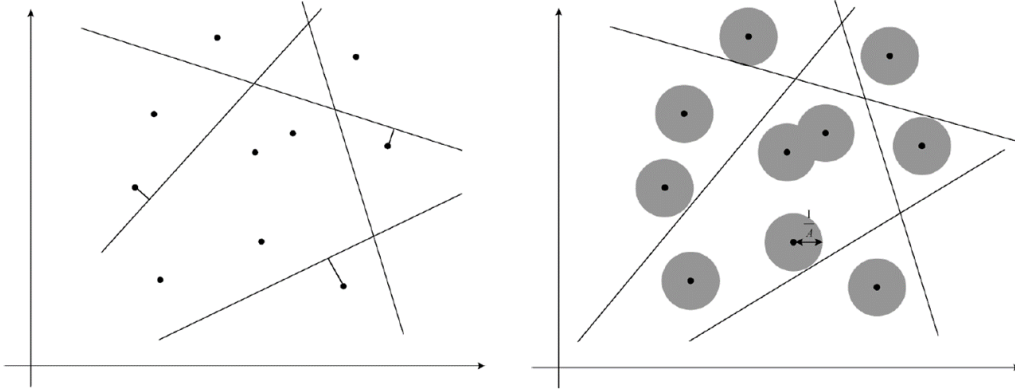


Figure 5.15 : Canonical hyperplanes and constraints (Gunn, 1998, pp. 6,7).

The bounding of the VC dimension of the canonical hyperplane set can be defined as,

$$h_{VC} \leq \min[R^2 A^2, n] + 1 \quad (5.15)$$

Here h_{VC} is the VC dimension, R is the radius of a hypersphere enclosing all the data points, and n is the dimension. To minimize (5.14), the upper bound of the VC dimension need to be minimized. It is possible to minimize (5.14) under the constraints of (5.8) by using Lagrangian.

$$\Phi(w, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i (y^i [\langle \mathbf{w}, \mathbf{x}^i \rangle + b] - 1) \quad (5.16)$$

Here α is the Lagrangian multiplier. According to Gunn (1998), the Lagrangian must be minimized as for w , b and maximized respecting $\alpha \geq 0$ (p. 8). To make the problem be solved easier, (5.16) can be transformed to dual problem, which is given by,

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \left(\min_{w, b} \Phi(w, b, \alpha) \right) \quad (5.17)$$

The minimum of the Lagrangian with respect to w and b can be defined as,

$$\frac{\partial \Phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (5.18a)$$

$$\frac{\partial \Phi}{\partial w} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i = \mathbf{w} \quad (5.18b)$$

By combining (5.17), (5.18a) and (5.18b), the primal problem can be transformed into dual problem, which can be described as,

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{k=1}^l \alpha_k \quad (5.19)$$

And the solution to (5.19) is,

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{k=1}^l \alpha_k \quad (5.20)$$

with following constraints:

$$\alpha_i \geq 0 \mid i = 1, \dots, l \mid \sum_{j=1}^l \alpha_j y_j = 0 \quad (5.21)$$

Equations (5.18a), (5.18b), and the constraints given in (5.21) are called as Karush-Kuhn-Tucker (KKT) conditions, which is introduced by Fletcher (2000). According to Burges (1998), “The KKT conditions are satisfied at the solution of any constrained optimization problem (convex or not), with any kind of constraints, provided that the intersection of the set of feasible directions with the set of descent directions coincides with the intersection of the set of feasible directions for linearized constraints with the set of descent directions.” (p. 10).

Applying the constraints given in (5.21) to (5.20) and solving yields the optimal separating hyperplane, which is given by,

$$w^* = \sum_{i=1}^l \alpha_i y_i x_i \quad (5.22)$$

$$b^* = -\frac{1}{2} \langle w^*, x_r + x_s \rangle \quad (5.23)$$

Here x_r and x_s are any support vector from each corresponding class.

In general, the classifiers can be described in two groups in terms of type, which are hard and soft margin classifiers (Boswell, 2002). The hard margin classifier is,

$$f(x) = \text{sgn}(\langle w^*, x \rangle + b) \quad (5.24)$$

And the soft margin classifier is,

$$f(x) = h_{VC}(\langle w^*, x \rangle + b) \text{ where } h(z) = \begin{cases} -1 & : z < -1 \\ z & : -1 \leq z \leq 1 \\ +1 & : z > 1 \end{cases} \quad (5.25)$$

It is stated by Gunn (1998) that applying soft margin is more proper than hard margin since soft margin generates a real value between -1 and 1 as an output when the classifier is objected within the margin, where no training data presents (p. 9).

Another KKT condition is described as,

$$\alpha_i (y^i [\langle w, x^i \rangle + b] - 1) = 0 \mid i = 1, \dots, l \quad (5.26)$$

Thus, only the points x^i that satisfy,

$$y^i[\langle w, x^i \rangle + b] = 1 \quad (5.27)$$

have non-zero Lagrange multipliers, which are called as Support Vectors (SV). If the data is linearly separable, as can be seen on Figure 5.13, all the SV will be located on the margin, thus the number of SV can be very small. In addition, the hyperplane is resolved by a small subset of the training data, as the other points could be excluded from the training set, which will be resulted in same hyperplane answer (Gunn, 1998, p. 9). In this situation, the following equation will be valid.

$$\|w\|^2 = \sum_{i=1}^l \alpha_i = \sum_{i \in SVs} \alpha_i = \sum_{i \in SVs} \sum_{j \in SVs} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (5.28)$$

Therefore, the bounds of the VC dimension of the classifier is,

$$h_{VC} \leq \min \left[R^2 \sum_{i \in SVs} , n \right] + 1 \quad (5.29)$$

5.2.2 The generalized optimal separating hyperplane

All the expressions above are valid for the case that the training data is linearly separable. However, this situation is not present all the time in general as illustrated in Figure 5.16.

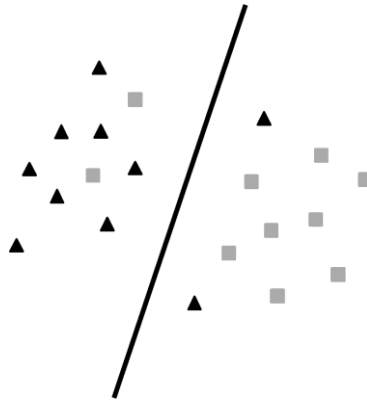


Figure 5.16 : Generalized optimal separating hyperplane.

There are two approaches to determine the generalized optimal separating hyperplane shown in Figure 5.16. First approach depends on the previous learning of the problem, and the second bases on estimating the noise in the data (Gunn, 1998, p. 10). To make the hyperplane separate the data correctly, an additional cost function

related to misclassification is inserted. To generalize the optimal separating hyperplane, a penalty function with non-negative variables that introduced by Cortes and Vapnik (1995) is implemented.

$$F(\xi) = \sum_i \xi_i \quad (5.30)$$

Here ξ_i are the slack variables, which is a measure of the misclassification errors. The optimization problem is now converted into minimizing the classification error and the bound of VC dimension of the classifier, simultaneously. Modifying the constraints in (5.8) for the non-separable case yields,

$$y^i[\langle \mathbf{w}, \mathbf{x}^i \rangle + b] \geq 1 - \xi_i \mid i = 1, \dots, l \mid \xi_i \geq 0 \quad (5.31)$$

Thus, the generalized optimal separating hyperplane can be determined by \mathbf{w} that minimizes the Lagrangian.

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \quad (5.32)$$

Here C is the complexity parameter, which is a user-defined value for adjusting margin maximization versus slack variable minimization. Choosing a high value of C will result in high penalty, while choosing a low value lets margin errors with possible misclassifications (Burges, 1998, p. 14). Naming of C is mentioned by Flach (2012) that if more margin errors is allowed, fewer SVs will be needed, so that C controls the complexity of the SVM, which can be found similar to the regularization of least-squares regression (p. 217).

The subject is finding solution to (5.32) under the constraints of (5.31), which can be performed via,

$$\Phi(\mathbf{w}, b, \alpha, \xi, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i - \sum_{i=1}^l \alpha_i (y^i [\mathbf{w}^T \mathbf{x}^i + b] - 1 + \xi_i) - \sum_{j=1}^l \beta_j \xi_j \quad (5.33)$$

Here α and β are the Lagrangian multipliers. The goal is to minimize the Lagrangian with respect to \mathbf{w} , b , \mathbf{x} and maximize it with respect to α and β . As mentioned before, Lagrangian duality lets (5.33) transform from primal problem to its dual problem, which is given by,

$$\max_{\alpha} W(\alpha, \beta) = \max_{\alpha, \beta} \left(\min_{w, b, \xi} \Phi(w, b, \alpha, \xi, \beta) \right) \quad (5.34)$$

By combining (5.32), (5.33), and (5.34), the dual problem becomes,

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{k=1}^l \alpha_k \quad (5.19)$$

And the solution to (5.19) is,

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{k=1}^l \alpha_k \quad (5.20)$$

with following constraints:

$$0 \leq \alpha_i \leq C \mid i = 1, \dots, l \mid \sum_{j=1}^l \alpha_j y_j = 0 \quad (5.35)$$

Additionally, Gunn (1998) stated that C must be selected in order to emphasize the identity of the noise in the training data (p. 12).

5.3 Kernel Functions

As explained before, in SVM, the optimal separating hyperplane is regulated in order to maximize the generalization ability. However, if the training data is linearly non-separable, the classifier may not have the capability of generalization despite the optimally determined hyperplane, hence the input space is mapped into a high-dimensional space that created by dot product, which is named as “feature space” (Abe, 2010, p. 31). The mapping of the input space into a high-dimensional feature space is illustrated in Figure 5.17.

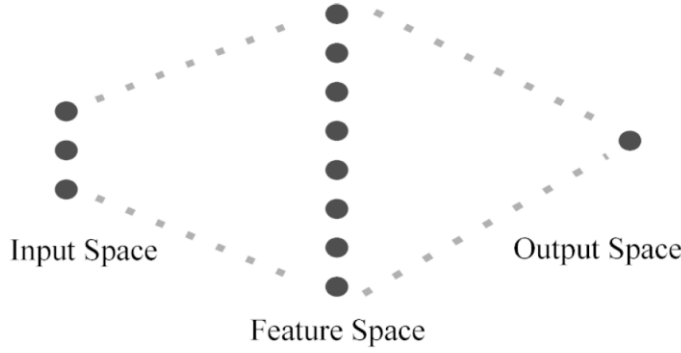


Figure 5.17 : Feature space illustration (Gunn, 1998, p. 14).

Kernel function's theory has a basis on Reproducing Kernel Hilbert Spaces (RKHS), which is introduced by several authors (Aronszajn, 1950; Girosi, 1997; Heckman, 1997; Wahba, 1990) and Mercer's Condition (Mercer, 1909). As stated above, kernel in input space is equal to an inner product in the feature space. Thus, kernel function is expressed as,

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \quad (5.36)$$

where,

$$\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_l(\mathbf{x}))^T \quad (5.37)$$

Here $K(\mathbf{x}, \mathbf{x}')$ is the kernel function, which performs linear or non-linear mapping of the input space into the feature space, \mathbf{x}' is unlabeled input, and l is the dimension number of the feature space.

There are several kernel functions such as linear, polynomial, Gaussian radial basis function, exponential radial basis function, multi-layer perceptron, Fourier series, splines, B splines, additive kernels, and tensor product (Gunn, 1998). In this thesis study, the linear and Gaussian radial basis function kernels are used for calculations. Thus, only these kernels' equations are given. For more details, please read the technical report of Gunn (1998).

5.3.1 Linear kernels

Linear kernels are used when the data to be classified is linearly separable in the input space (Abe, 2010, p. 34). In this situation, it is not necessary to map the input space into the feature space. Linear kernels are expressed as,

$$K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}' \quad (5.38)$$

5.3.2 Gaussian radial basis function kernels

Another kernel used in this thesis study is the radial basis functions (RBF) in Gaussian form, which can be expressed as,

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (5.39)$$

Here γ is a positive user-defined parameter, which controls the radius.

5.4 Support Vector Regression

The concept explained above is about SVM that classifies data inputs into different classes. However, the problem sometimes may require less rigid and more possible results. Support Vector Regression (SVR) is a kind of implementation of SVM, which generates a real number as an output (Flach, 2012).

The main principle of SVR is same as SVM for classification with several minor differences. Basically, SVM is applicable to regression problems via importing an alternative loss function, which was introduced by Smola (1996). Different possible loss functions are shown in Figure 5.18.

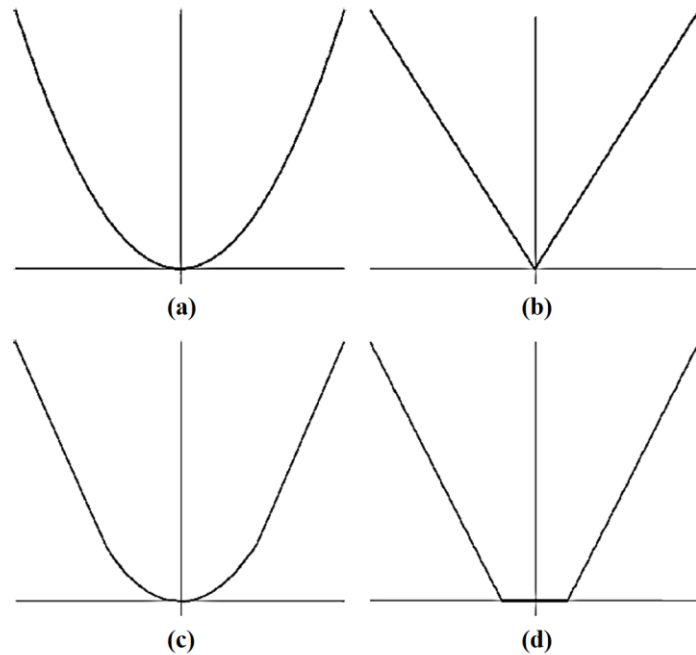


Figure 5.18 : Loss functions (Gunn, 1998, p. 29).

In Figure 5.18, (a) is the quadratic loss function, which is related to the generic least-squares error criteria. (b) is the Laplacian type of loss function, which has less sensitivity to deviation rather than the quadratic loss function. (c) is the loss function proposed by Huber (1964), which includes optimal features when the data distribution is unknown. However, as mentioned by Gunn (1998), these loss functions does not produce any sparseness in support vectors (p. 30). Thus, the loss function (d) introduced by Vapnik (1999), which has a similar approach with Huber's loss function is used to obtain support vectors.

5.4.1 ε -insensitive regression

Assuming an approximation problem including the set of data in (5.40).

$$\mathcal{D} = \{(x^1, y^1), \dots, (x^l, y^l)\} \mid x \in \mathbb{R}^n \mid y \in \mathbb{R} \quad (5.40)$$

Using a linear function such as mentioned before,

$$f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (5.5)$$

The optimal regression function can be found by minimizing (5.41).

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i (\xi_i^-, \xi_i^+) \quad (5.41)$$

Here ξ^- and ξ^+ are the lower and upper constraints of slack variables, respectively. The loss function of (d) in Figure 5.18, in other words ε -insensitive loss function, can be expressed as,

$$L_\varepsilon(y) = \begin{cases} 0, & |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon, & \text{otherwise} \end{cases} \quad (5.42)$$

The illustration of ε -insensitive loss function and slack variables are shown in Figure 5.19.

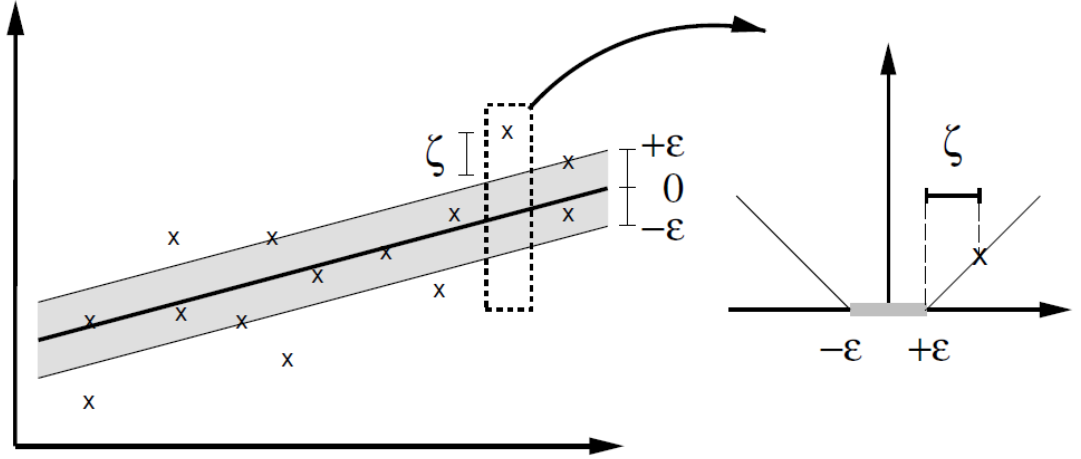


Figure 5.19 : ε -insensitive loss function and slack variables (Smola and Schölkopf, 2003, p. 2).

The solution to the problem in (5.42) is,

$$\bar{\alpha}, \bar{\alpha}^* = \arg \min_{\alpha, \alpha^*} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i + \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varepsilon \quad (5.43)$$

with following constraints:

$$0 \leq \alpha_i, \alpha_i^* \leq C \mid i = 1, \dots, l \mid \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (5.44)$$

Solving (5.43) with the constraints of (5.44) yields the Lagrange multipliers, α and α^* . Thus, the function of regression becomes,

$$\bar{\mathbf{w}} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (5.45)$$

$$\bar{b} = -\frac{1}{2} \langle \bar{\mathbf{w}}, (\mathbf{x}_r + \mathbf{x}_s) \rangle \quad (5.46)$$

The solution satisfies the KKT conditions, which are,

$$\bar{\alpha}_i \bar{\alpha}_i^* = 0 \mid i = 1, \dots, l \quad (5.47)$$

Hence, support vectors are non-zero Lagrange multipliers. If ε is set at 0, the optimization problem becomes,

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^l \beta_i y_i \quad (5.48)$$

with following constraints:

$$-C \leq \beta_i \leq C \mid i = 1, \dots, l \mid \sum_{i=1}^l \beta_i = 0 \quad (5.49)$$

Thus, the function of regression becomes,

$$\bar{\mathbf{w}} = \sum_{i=1}^l \beta_i \mathbf{x}_i \quad (5.50)$$

$$\bar{b} = -\frac{1}{2} \langle \bar{\mathbf{w}}, (\mathbf{x}_r + \mathbf{x}_s) \rangle \quad (5.46)$$

5.4.2 ν regression

The summary of ε -insensitive regression can be described as follows. For determining the function of (5.5) with the data given as (5.40), each \mathbf{x}_i has an error of ε , while all the other values above ε are slack variables, ξ_i , that correspond to the cost of the main function with the regularization constant, C (Schölkopf et al, 1998). However, it is quite difficult to determine the optimum value of ε (Abe, 2010, p.431). One of the approaches to solve this optimization problem was proposed by Smola et al. (1998) and Jeng and Chuang (2002), which assumes the estimated optimum value is proportional to the standard deviation of the noise in the data. Another approach is introduced by Schölkopf et al. (1998) which modifies the model by optimizing it during training phase with a controller parameter, ν .

In ν -regression, the amount of ε is adjusted contrary to the complexity of the model and slack variables with a constant, ν , by minimizing following equation (Schölkopf et al, 1998).

$$\Phi(\mathbf{w}, \xi, \varepsilon) = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\nu \varepsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \right) \quad (5.47)$$

with following constraints:

$$(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \varepsilon + \xi_i \quad (5.48)$$

$$y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \quad (5.49)$$

$$\xi \geq 0 \mid \varepsilon \geq 0 \quad (5.50)$$

Here ξ_i and ξ_i^* are non-negative slack variables. The solution is,

$$\begin{aligned} \Phi(\mathbf{w}, b, \alpha, \xi, \beta, \varepsilon, \eta) = \\ \frac{1}{2} \|\mathbf{w}\|^2 + C\nu\varepsilon + \frac{C}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (\xi_i + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b + \varepsilon) \\ - \sum_{i=1}^l \alpha_i^* (\xi_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b + \varepsilon) - \beta\varepsilon - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (5.51)$$

Here α , β and η are non-negative Lagrange multipliers. In order to minimize (5.47) the saddle point of (5.51) needs to be found by minimizing primal variables, \mathbf{w} , ε , b , ξ and maximizing dual variables, α , β , η (Schölkopf et al, 1998). Setting the derivatives of primal variables to zero yields,

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \mathbf{x}_i \quad (5.52)$$

and,

$$C\nu - \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \beta = 0 \quad (5.53)$$

As a summary, ν -SVR optimization problem becomes as maximizing,

$$W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle (\alpha_j - \alpha_j^*) - \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \quad (5.54)$$

with the following constraints:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (5.55)$$

$$0 \leq \alpha \leq \frac{C}{l} \quad (5.56)$$

$$\sum_{i=1}^l (\alpha_i + \alpha_i^*) = Cv \quad (5.57)$$

5.5 Cross Validation and Overfitting

As mentioned before, the main concept of the machine learning is to train a learning model on a training data set, and apply this model to a new data set to make predictions. The focus is to maximize the predicting accuracy on the new data set. Achieving maximum accuracy on the training data set is not necessary. However, in this situation, there is a risk in terms of fitting the noise in the data to learning curve by memorizing training data rather than establishing a general predictive bias, which is called as “overfitting” (Dietterich, 1995, p. 326). An example of the overfitting problem is shown in Figure 5.20.

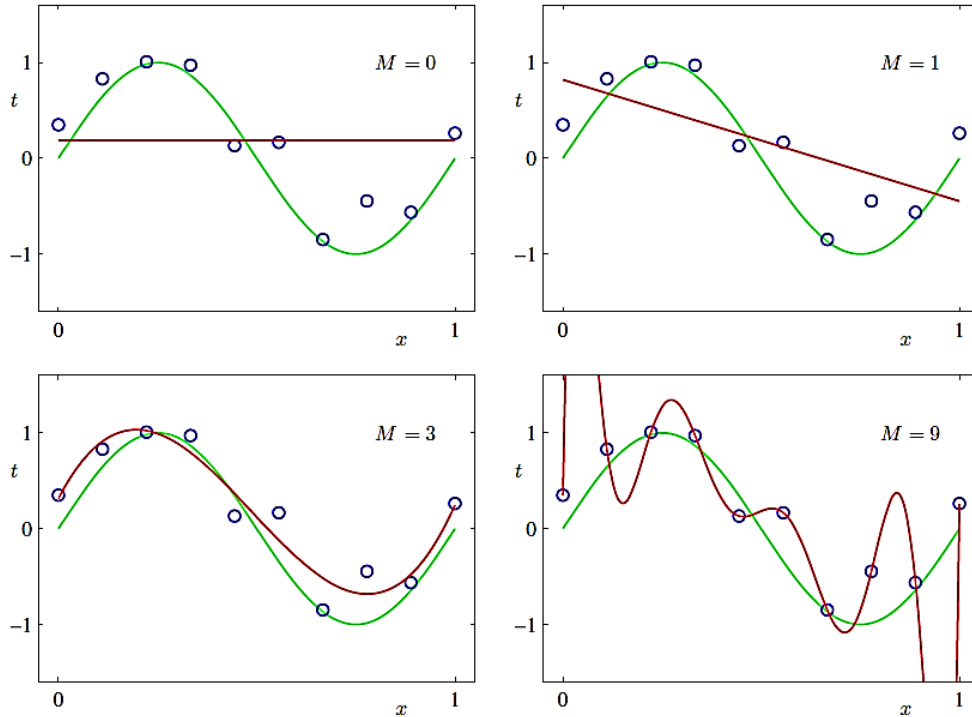


Figure 5.20 : Overfitting example (Bishop, 2006, p. 7).

In Figure 5.20, M represents the order of the polynomial. The green curve is the true function that used to generate the data, which is not a polynomial. The data in this example has a uniform distribution in x -axis, while having noise in y -axis. As seen on the figure, when the order of the polynomial increased, the curve fits the training data in excellent way. Moreover, the curve crosses each data point with zero error. However, there is an extreme oscillation when order of the polynomial is nine, and the representation of the original function is quite poor. This phenomenon is known as overfitting.

As depending on the order of the polynomial, the overfitting problem can depend on the number of input data as well. An example to this kind of a problem is shown in Figure 5.21.

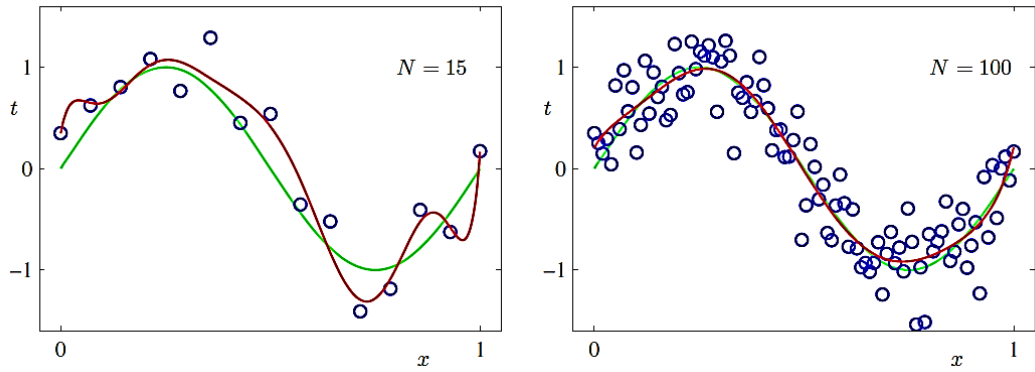


Figure 5.21 : Reducing overfitting (Bishop, 2006, p. 9).

Here N represents the amount of input data points. It can be seen from Figure 5.21 that as the number of input data increases the model becomes more complex to fit the curve in a more meaningful way. According to Bishop (2006), “One rough heuristic that is sometimes advocated is that the number of data points should be no less than some multiple (say 5 or 10) of the number of adaptive parameters in the model.” (p. 9).

In order to overcome the overfitting problem, a kind of a statistical analysis called “cross validation” is need to be applied to the training data set. Cross validation is a validation technique to determine the assessment of generalizing the statistical results on an independent data set, different from test data set (Flach, 2012). Thus, if a learning algorithm is evaluated, rather than a given model, it is needed to create a new data set separate from the test data. If the data set has limitations in terms of

input number, the cross validation process is applied by using two common approaches.

5.5.1 K-fold cross validation

First, the data is randomly divided into k folds. One of the folds is set separately for testing purpose. The model is trained on the remaining $k - 1$ folds and evaluated on the test fold for k folds until each fold has been used for testing once (Flach, 2012, p. 349). The most common setting for k -fold cross validation is to set k to 10, which is called as “10-fold cross validation”.

5.5.2 Leave-one-out cross validation

A second approach to k -fold cross validation is setting k to n , and training on all training inputs except one for n times. This process is called as “leave-one-out cross validation”. The statistical meaning of this kind of cross validation is estimating the accuracy of each fold as 0 or 1, while getting a distribution similar to normal distribution due to the central limit theorem by averaging n of the folds (Flach, 2012, p. 350).

5.6 Toolboxes

Solving all the optimization problems explained above analytically for the data sets containing huge number of inputs is not quite possible. Therefore, several toolboxes have been introduced for machine learning applications. The programming language, the libraries and other toolboxes used in this thesis study is licensed under GNU General Public License.

5.6.1 R programming language

R is a statistical console suite created as a free-licensed implementation of the S programming language developed by Rick Becker, John Chambers and Alan Wilks at Bell Laboratories. In recent times, R become very popular thanks to the developing methods for interactive data analysis (Venables and Smith, 2014, p. 2). The most useful property of the R programming language is to be expandable by a large collection of external libraries called packages.

5.6.2 LIBSVM

LIBSVM is a library have been developing by Chang and Lin (2012) since 2000 for the usage of SVM. LIBSVM is the most common and widely used SVM library (Chang and Lin, 2011, p. 1). LIBSVM works in two steps. In the first step, a training data is set to produce a model. In the second step, the model is used on a test set to make predictions.

5.6.3 e1071 package

e1071 is an award-winning C++ implementation of the LIBSVM that can be imported in R programming language as a package (Meyer, 2014). e1071 performs most of the SVM features, such as C and ν -classification, novelty detection, ε and ν -regression. e1071 also contains linear, polynomial, RBF, sigmoid kernels, formula interface, and k -fold cross validation.

A simple code of the program written in R programming language using e1071 package for the calculations of this thesis study is given in Syntax 1. Full code with detailed explanations is available in Appendix B.

Syntax 1

```
library(xlsx)
library(e1071)
dataset <- read.xlsx("data.xls",1)
dataset.train <- dataset[dataset$subset=="train",2:ncol(dataset)]
dataset.test <- dataset[dataset$subset=="test",2:ncol(dataset)]
def.pred <- mean(dataset.train$x1)
def.rss <- sum((dataset.test$x1-def.pred)^2)
reg <- lm(x1 ~., data = dataset.train)
reg.pred <- predict(reg,newdata = dataset.test)
reg.rss <- sum((dataset.test$x1-reg.pred)^2)
obj <- tune.svm(x1 ~.,data = dataset.train, scale = T, type = "eps-regression",
  kernel = "linear", cost = seq(from=0.005,to=50.0,by=0.005),epsilon=0.1,
  tolerance=0.001, shrinking=T, fitted=T)
k <- which.min(obj$performances[,3])
c <- obj$performances[,1][k]
RMSE <- obj$performances[,3]
Cost <- obj$performances[,1]
plot(Cost,RMSE,type="l")
epsilon.svr <- svm(x1 ~.,data = dataset.train, scale = T, type = "eps-
  regression", kernel = "linear", cost = c,
  epsilon=0.1,tolerance=0.001,shrinking=T,fitted=T)
esvr2.pred <- predict(epsilon.svr,newdata = dataset.test)
esvr2.rss <- sum((dataset.test$x1-esvr2.pred)^2)
```

5.7 Statistical Comparison Criteria

To compare the results gathered from either multiple regression analysis or all the SVR implementations explained above, several statistical comparison criteria were used. The criteria is the basic statistical components being used in most of the

statistical analysis applications, which are residual sum of squares, R^2 , pseudo- R^2 , root mean squared error, and p-value.

5.7.1 Residual sum of squares

As explained in Chapter 4, residual is the difference between predicted and observed value of a function. By modifying the equation form of residual, (4.15a), Residual sum of squares (RSS) can be expressed as,

$$RSS = \sum_i (y_i - \hat{y}_i)^2 \quad (5.58)$$

As a summary, RSS gives the measure of the variation of the observed values around the fitted curve.

5.7.2 R^2

R^2 , or coefficient of determination, is a measure of the perfectness of the fitted curve. According to Fonticella (n.d.), "...the R^2 statistic measures how significantly the slope of the fitted line differs from zero, which is not the same as a good fit." (p. 56).

As an equation form, R^2 can be described as,

$$R^2 = \frac{RSS}{TSS} \quad (5.59)$$

where,

$$TSS = \sum_i (y_i - \bar{y})^2 \quad (5.60)$$

Here TSS stands for Total sum of squares, which is a measure of the variation of the observed values around the mean.

An important drawback of the R^2 is the disability of decrease when the model gains additional informative variables, although these variables have no relevance with the function (Dufour, 2011, p. 5). Thus, the approach that tends to maximize R^2 can be deceptive. To overcome this handicap, adjusted R^2 is introduced to be used instead of R^2 , since increasing number of unnecessary variables creates a penalty on the R^2 . Adjusted R^2 is expressed as,

$$R_{adj}^2 = 1 - \left(\frac{RSS}{TSS} \right) \left(\frac{n-1}{n-p} \right) \quad (5.61)$$

Here n is the number of samples, and p is the number of regressors.

5.7.3 Pseudo- R^2

Pseudo- R^2 is an alternative form used for determining the validity of a model on a reference situation (Williams, 2015, p. 3). It is recommended to use Pseudo- R^2 to overcome any overfitting problem if the number of data inputs are limited (Mittlböck and Heinzl, 2004).

As explained in the previous chapters, according to least square principle, the default model is simply the mean of the dependent variable computed on the training sample. Thus, the RSS of the default model is,

$$RSS_{default} = \sum_i (y_i - \bar{y}_{train})^2 \quad (5.62)$$

Hence, the Pseudo- R^2 criterion can be defined as,

$$Pseudo - R^2 = 1 - \frac{RSS}{RSS_{default}} \quad (5.63)$$

The statistical meaning of (5.63) can be described as following.

- $Pseudo - R^2 = 1 \Rightarrow$ Perfect model
- $Pseudo - R^2 = 0 \Rightarrow$ Not better than default model
- $Pseudo - R^2 < 0 \Rightarrow$ Worse than default model

5.7.4 Root mean squared error

Root mean squared error (RMSE) is a measure of verification, which standardize the variables between matching points of predictions and observations (Barnston, 1992, p. 700). RMSE can be expressed as,

$$RMSE = \sqrt{\frac{1}{n} \sum_i e_i^2} \quad (5.64)$$

6. RESULTS AND DISCUSSIONS

Interpreting the results, the datasets in Appendix A are employed in the calculations. The data set given in Table A.1 obtained from Bourgoyne Jr. and Young Jr. (1974) is used to calculate the optimization of ROP. Assignment of the dependent and independent variables, and the steps of multiple regression solution are given in Appendix A, in details.

The main goal of the calculations are to determine the applicability of SVR for ROP optimization when there are insufficient data inputs for multiple regression analysis. To achieve a general idea, several case scenarios are also designed. The calculation sequence constructed as the following cases.

Case 1: Training top 16 depth data are used for testing deeper data points.

Case 2: Training bottom 16 depth data are used for testing shallower data points.

Case 3: Training odd-numbered data points are used for testing remaining.

Case 4: Training even-numbered data points are used for testing remaining.

Case 5: Training several 24 data points are used for testing remaining.

Case 6: Training all data points are used for testing several data points.

Case 7: Training different data set are used for testing all data points (for comparison).

In testing parts of the calculations, different variations of data combinations are used as testing sets. These combinations are categorized under different scenarios and the results are discussed separately.

The cost parameter, C , is calculated via 10-fold cross validation within a given interval. The cost value gives minimum root mean squared error (RMSE) is selected as the optimum cost parameter for the corresponding data set and SVR method. The other SVR parameters are selected through the best performance of the current method.

6.1 Case 1: Training Top 16 Depth Interval

In Case 1, first 16 data points of Table A.1 are used as training data set. The coefficients of MR analysis for this training data are given in Table 6.1.

Table 6.1 : Multiple regression coefficients of Case 1.

a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
3,923E+00	3,880E-05	1,262E-04	4,354E-05	5,732E-01	5,913E-01	4,619E-01	-1,665E-01

6.1.1 Testing #17-18-19-20

In this testing scenario, the ROP values of the inputs numbered #17-18-19-20 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.1.

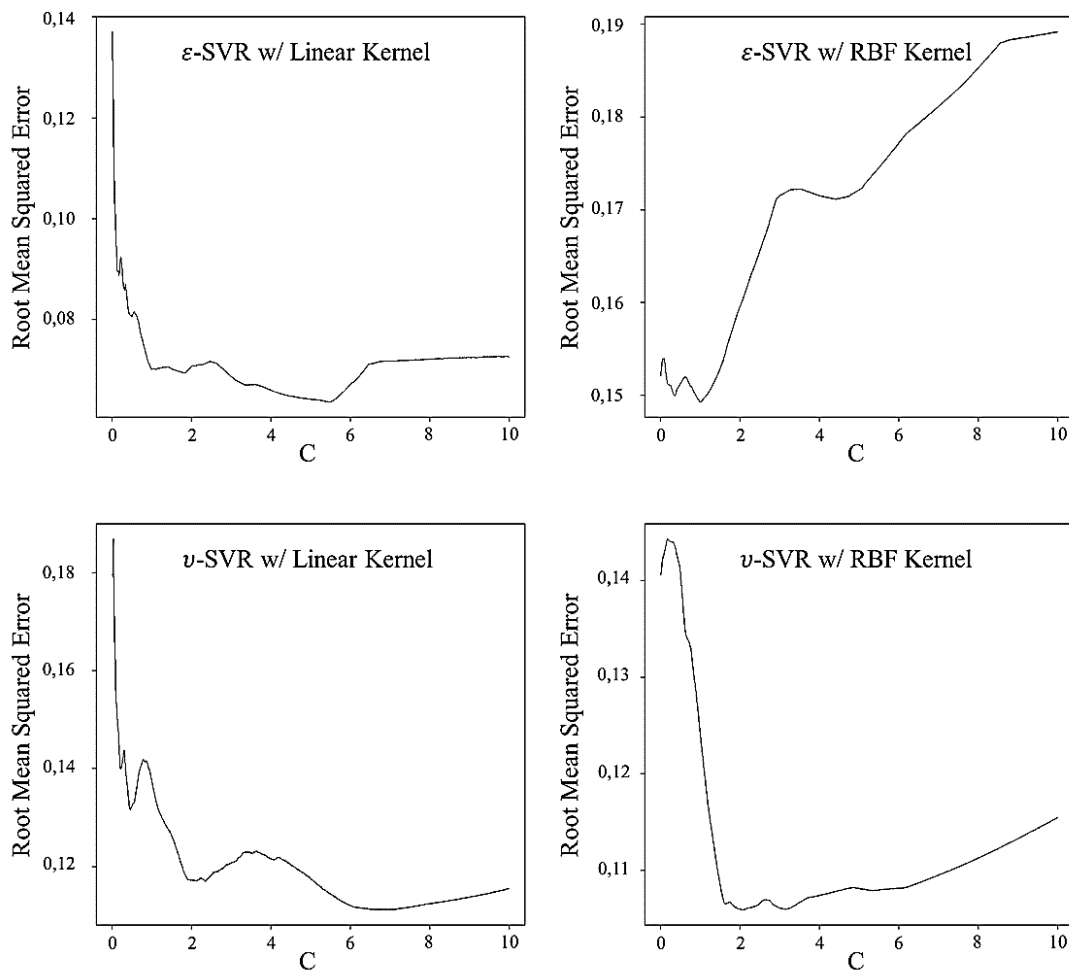


Figure 6.1 : Cost graphics of Case 1 for testing #17-18-19-20.

The basic statistical analysis of the whole data set is given in Table 6.2.

Table 6.2 : Statistics of the data set in Case 1 for testing #17-18-19-20.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,850	-1480	1758	-10844	-0,755	-0,218	-0,408	-0,074
Median	2,760	-1268	312	-7529	-0,730	-0,255	-0,390	-0,024
Variance	0,257	1595326	4567737	85108973	0,177	0,044	0,051	0,097

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.3.

Table 6.3 : Statistical results of MR in Case 1 for learning and testing #17-18-19-20.

RSS _{default}	R ²	Adjusted R ²	p-value
3,16	0,84	0,71	0,01

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.4.

Table 6.4 : Statistical results of the methods in Case 1 for testing #17-18-19-20.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,80	0,46	1,60	1,15	0,76
Pseudo-R ²	0,75	0,85	0,50	0,64	0,76
C		5,485	1	6,64	2,04
# of SV		13	15	11	14
CV Error		0,06	0,15	0,11	0,11
RMSE	0,45	0,34	0,63	0,54	0,44

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.5. The ROP values are in unit of [ft/hr].

Table 6.5 : ROP predictions in Case 1 for #17-18-19-20.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
17	42,1	43,0	19,6	39,8	24,1	42,7
18	38,5	38,9	20,4	36,4	25,9	38,6
19	33,1	32,8	20,9	29,8	27,2	43,4
20	29,3	23,2	15,6	34,0	16,0	12,5

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.2.

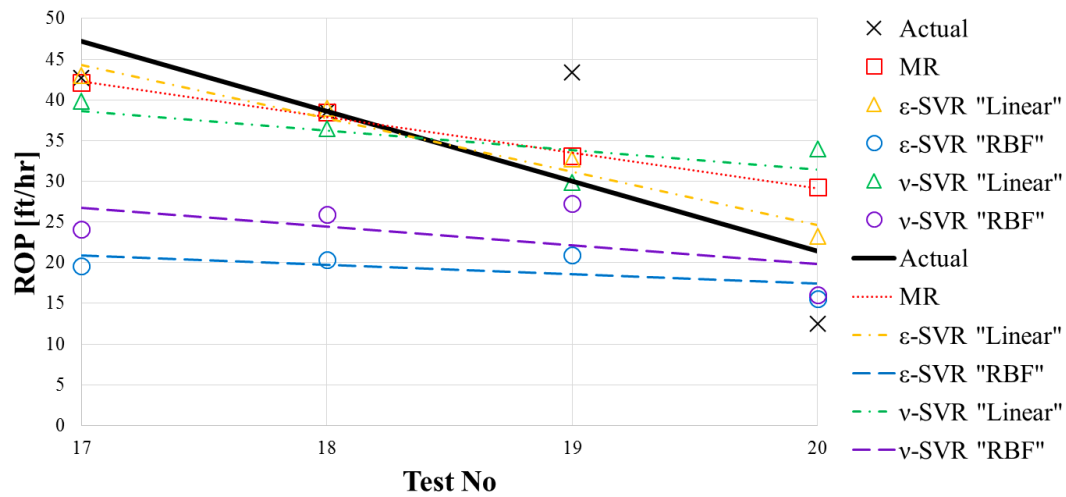


Figure 6.2 : ROP prediction trends in Case 1 for testing #17-18-19-20.

The comparison of each prediction with actual ROP is plotted in Figure 6.3.

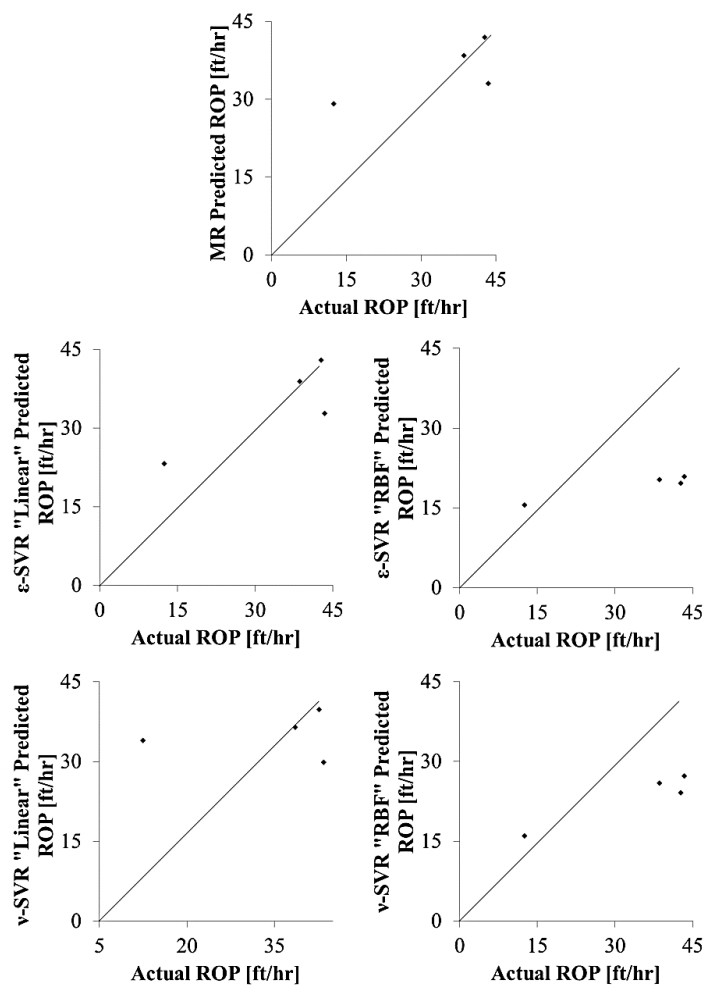


Figure 6.3 : ROP prediction comparison in Case 1 for testing #27-28-29-30.

It can be seen from Figure 6.2 and 6.3 that the best predictive method in Case 1 for testing #17-18-19-20 is ε -SVR with linear kernel, since it has the best correlation between predicted and actual ROP values. Additionally, ε -SVR with linear kernel has the minimum errors and maximum Pseudo- R^2 ratio among the other methods, as seen in the results.

6.1.2 Testing #22-23-24-25

In this testing scenario, the ROP values of the inputs numbered #22-23-24-25 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.4.

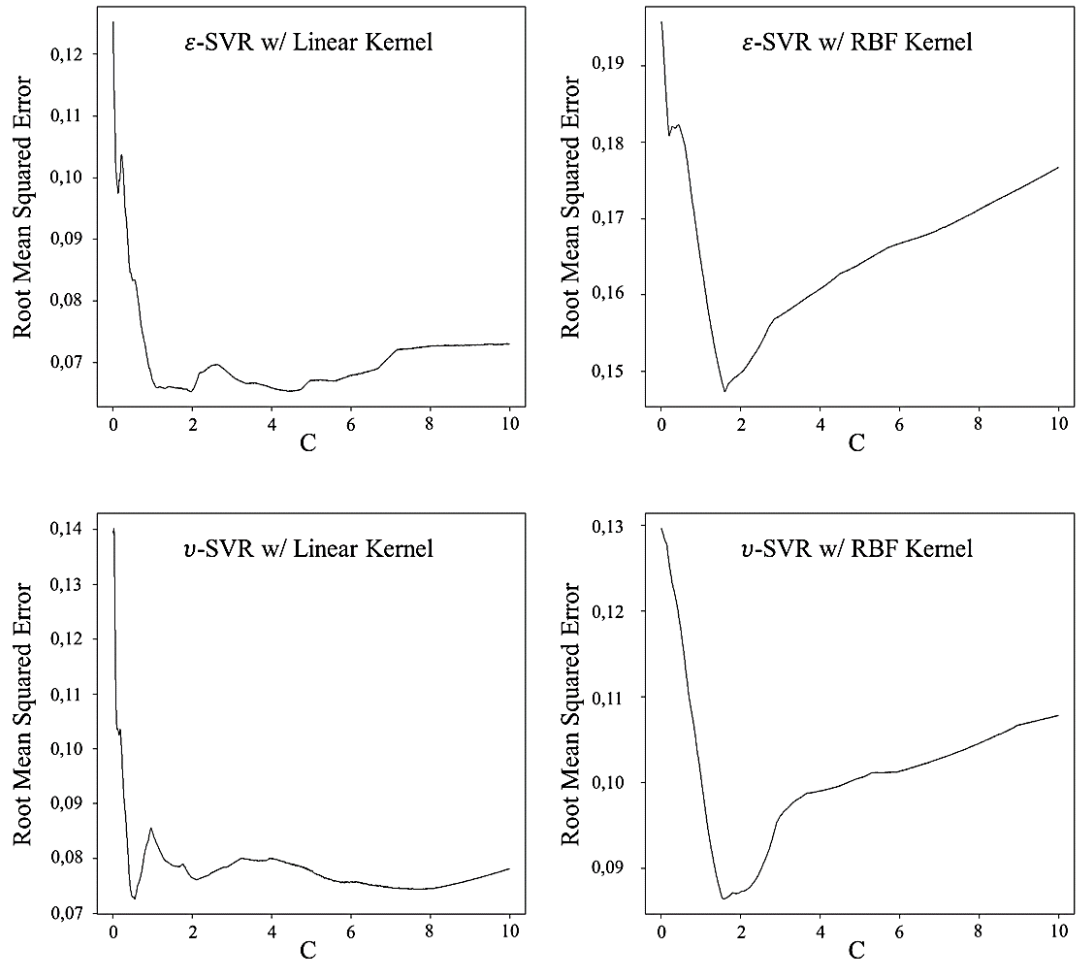


Figure 6.4 : Cost graphics of Case 1 for testing #22-23-24-25.

The basic statistical analysis of the whole data set is given in Table 6.6.

Table 6.6 : Statistics of the data set in Case 1 for testing #22-23-24-25.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,774	-1797	1862	-11399	-0,809	-0,224	-0,389	-0,089
Median	2,769	-1268	312	-7474	-0,738	-0,255	-0,355	-0,037
Variance	0,137	3210229	5332154	83943489	0,131	0,046	0,051	0,050

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.7.

Table 6.7 : Statistical results of MR in Case 1 for learning and testing #22-23-24-25.

RSS _{default}	R ²	Adjusted R ²	p-value
0,55	0,84	0,71	0,01

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.8.

Table 6.8 : Statistical results of the methods in Case 1 for testing #22-23-24-25.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,37	0,07	0,26	0,36	0,27
Pseudo-R ²	0,33	0,87	0,52	0,34	0,52
C		1,97	1,6	0,55	1,58
# of SV		14	14	12	14
CV Error		0,07	0,12	0,07	0,09
RMSE	0,30	0,13	0,26	0,30	0,26

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.9. The ROP values are in unit of [ft/hr].

Table 6.9 : ROP predictions in Case 1 for #22-23-24-25.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
22	25,1	19,9	15,8	16,6	15,8	19,0
23	24,4	19,5	16,1	18,2	16,1	18,7
24	31,1	16,4	15,9	11,8	15,9	20,2
25	32,1	31,5	18,4	34,1	18,3	27,1

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.5.

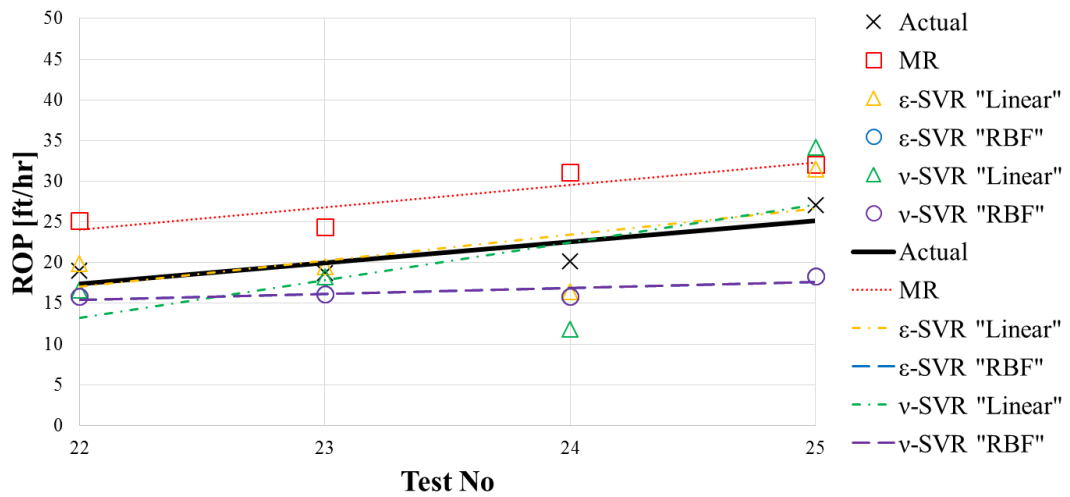


Figure 6.5 : ROP prediction trends in Case 1 for testing #22-23-24-25.

The comparison of each prediction with actual ROP is plotted in Figure 6.6.

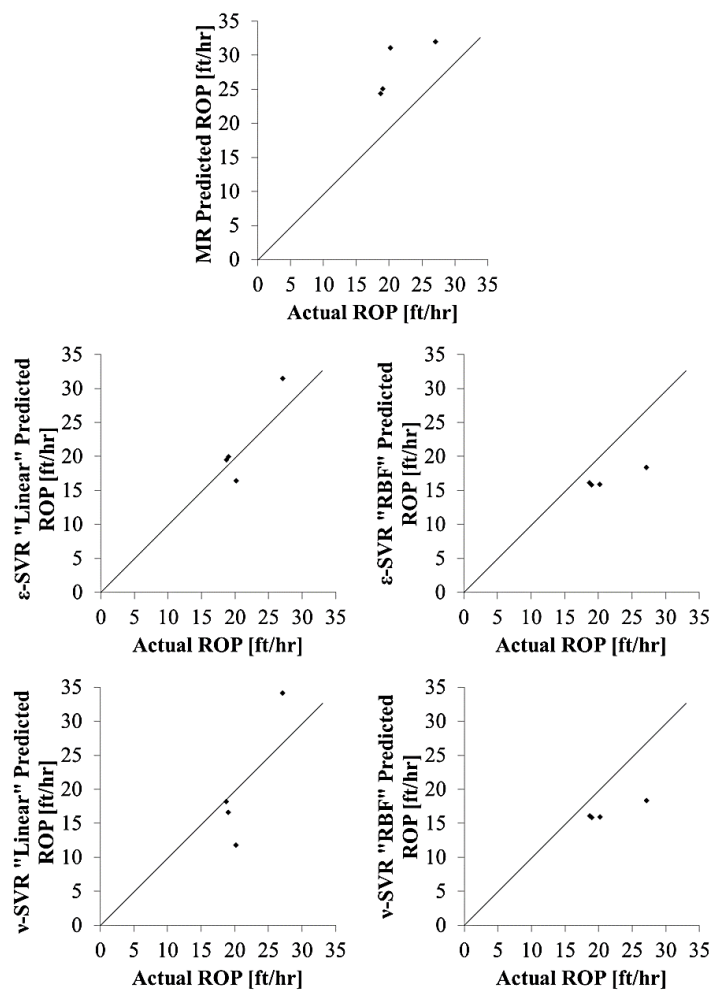


Figure 6.6 : ROP prediction comparison in Case 1 for testing #22-23-24-25.

Similarly to first scenario, as seen on Figure 6.5 and 6.6, the highest correlation is in ε -SVR with linear kernel. Moreover, ε -SVR with linear kernel has all the minimum errors and maximum Pseudo- R^2 value. In addition, the best approaches are seemed to be linear kernel methods, because RBF kernels do not give representative results for good prediction.

6.1.3 Testing #27-28-29-30

In this testing scenario, the ROP values of the inputs numbered #27-28-29-30 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.7.

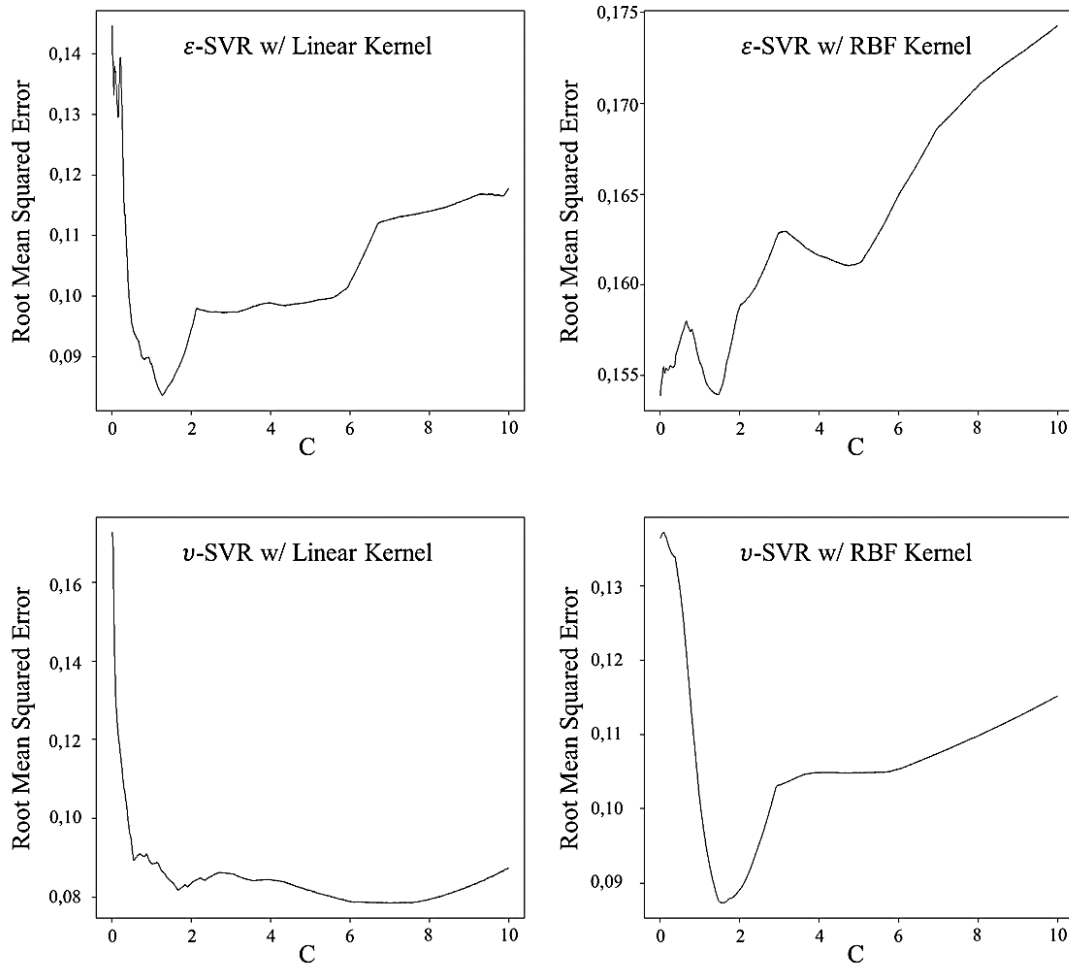


Figure 6.7 : Cost graphics of Case 1 for testing #27-28-29-30.

The basic statistical analysis of the whole data set is given in Table 6.10.

Table 6.10 : Statistics of the data set in Case 1 for testing #27-28-29-30.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,668	-2317	2034	-13085	-0,713	-0,235	-0,413	-0,133
Median	2,670	-1268	312	-8226	-0,669	-0,236	-0,390	-0,054
Variance	0,126	8250056	6815509	95070470	0,202	0,050	0,050	0,048

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.11.

Table 6.11 : Statistical results of MR in Case 1 for learning and testing #27-28-29-30.

RSS _{default}	R ²	Adjusted R ²	p-value
0,30	0,84	0,71	0,01

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.12.

Table 6.12 : Statistical results of the methods in Case 1 for testing #27-28-29-30.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	2,07	0,12	0,32	0,23	0,40
Pseudo-R ²	-6,02	0,60	-0,08	0,21	-0,36
C		1,27	0,005	7,035	1,565
# of SV		14	14	12	14
CV Error		0,08	0,15	0,08	0,09
RMSE	0,72	0,17	0,28	0,24	0,32

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.13. The ROP values are in unit of [ft/hr].

Table 6.13 : ROP predictions in Case 1 for #27-28-29-30.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
27	21,3	10,8	15,2	14,2	15,9	12,6
28	34,0	12,2	15,2	22,4	15,9	14,9
29	20,9	11,7	15,2	11,3	15,9	13,8
30	23,7	7,7	15,2	10,1	15,9	9,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.8.

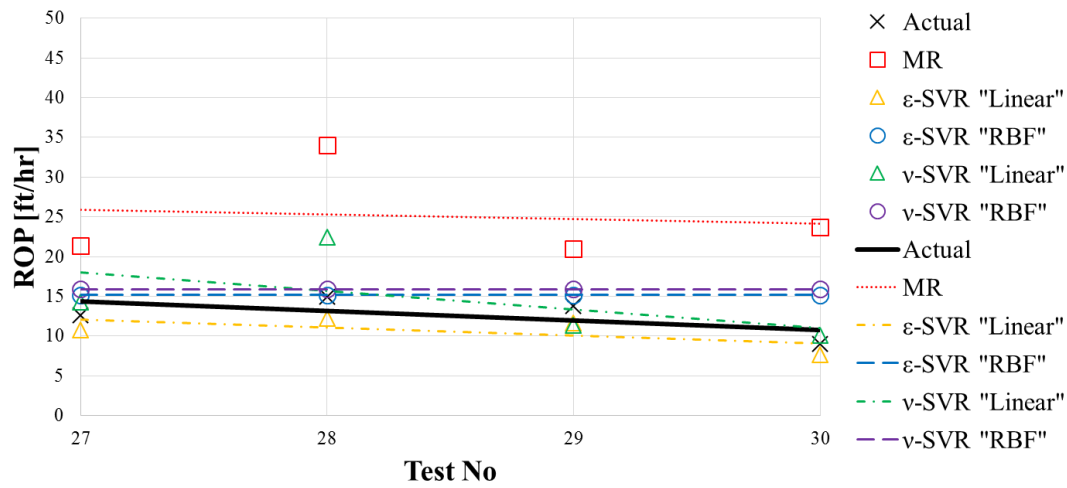


Figure 6.8 : ROP prediction trends in Case 1 for testing #27-28-29-30.

The comparison of each prediction with actual ROP is plotted in Figure 6.9.

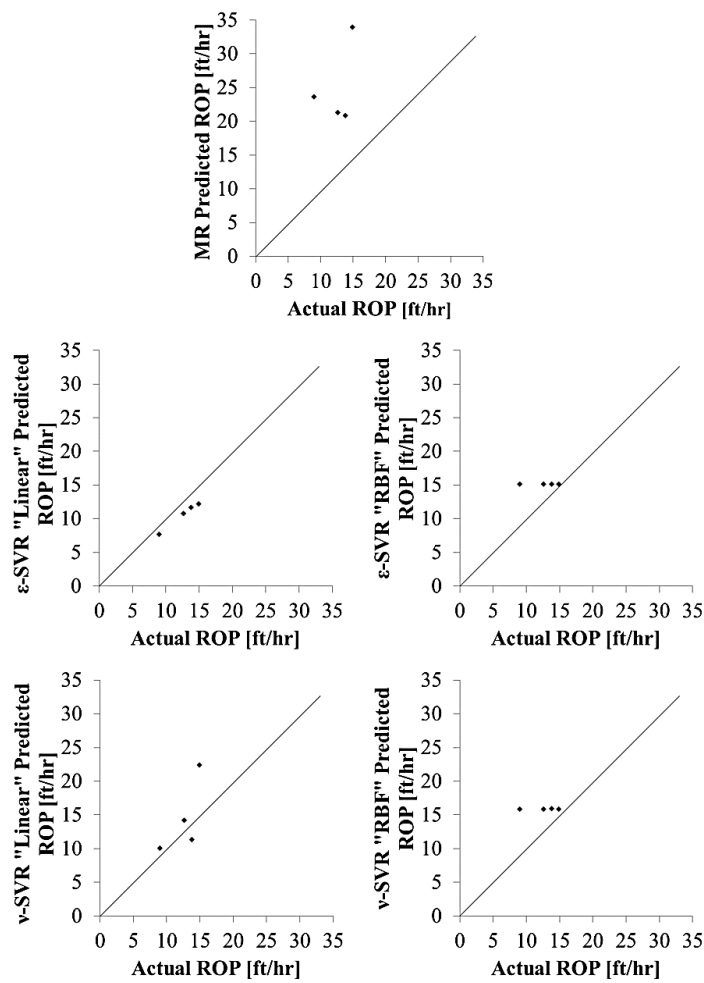


Figure 6.9 : ROP prediction comparison in Case 1 for testing #27-28-29-30.

Similarly to first and second scenario, the best prediction is in ε -SVR with linear kernel. Furthermore, all the minimum errors and maximum Pseudo- R^2 value are in ε -SVR with linear kernel. Additionally, the best predictors are seemed to be linear kernel methods as in the second scenario. It is quite obvious that RBF kernels do not represent meaningful results because of the visualisation of data.

6.1.4 Testing #17-21-26-30

In this testing scenario, the ROP values of the inputs numbered #17-21-26-30 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.10.

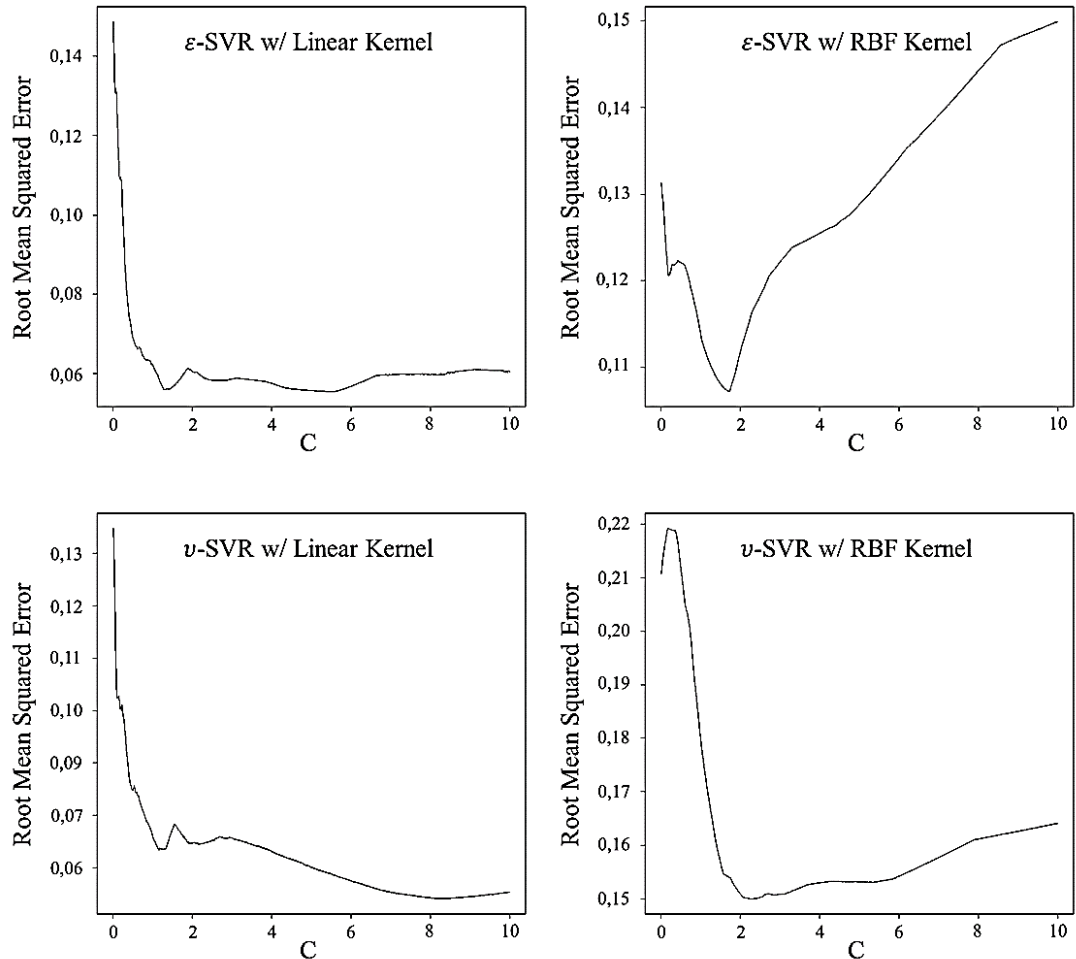


Figure 6.10 : Cost graphics of Case 1 for testing #17-21-26-30.

The basic statistical analysis of the whole data set is given in Table 6.14.

Table 6.14 : Statistics of the data set in Case 1 for testing #17-21-26-30.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,750	-1976	1912	-7529	-0,730	-0,255	-0,390	-0,037
Median	2,747	-1268	312	-11806	-0,783	-0,231	-0,408	-0,148
Variance	0,188	6113043	5869719	90241467	0,189	0,048	0,066	0,149

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.15.

Table 6.15 : Statistical results of MR in Case 1 for learning and testing #17-21-26-30.

RSS _{default}	R ²	Adjusted R ²	p-value
1,48	0,84	0,71	0,01

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.16.

Table 6.16 : Statistical results of the methods in Case 1 for testing #17-21-26-30.

	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	1,12	0,004	0,81	0,22	0,70
Pseudo-R ²	0,24	0,998	0,45	0,85	0,53
C		5,54	1,72	8,225	2,28
# of SV		13	13	11	15
CV Error		0,06	0,11	0,06	0,15
RMSE	0,53	0,03	0,45	0,23	0,42

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.17. The ROP values are in unit of [ft/hr].

Table 6.17 : ROP predictions in Case 1 for #17-21-26-30.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
17	42,1	43,1	22,8	41,2	25,3	42,7
21	23,9	20,2	15,9	33,2	16,1	21,1
26	22,4	14,7	16,1	14,3	16,4	14,8
30	23,7	8,7	15,9	10,1	16,1	9,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.11.

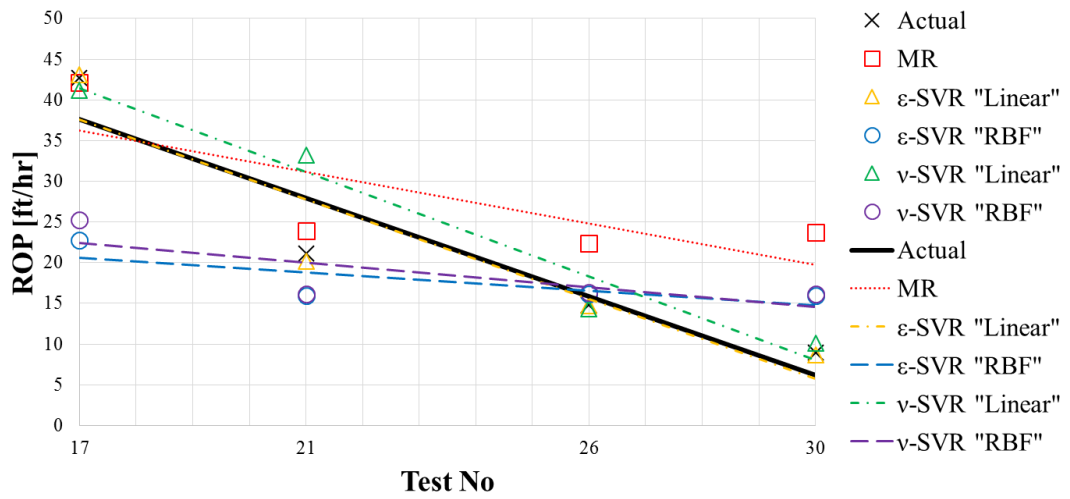


Figure 6.11 : ROP prediction trends in Case 1 for testing #17-21-26-30.

The comparison of each prediction with actual ROP is plotted in Figure 6.12.

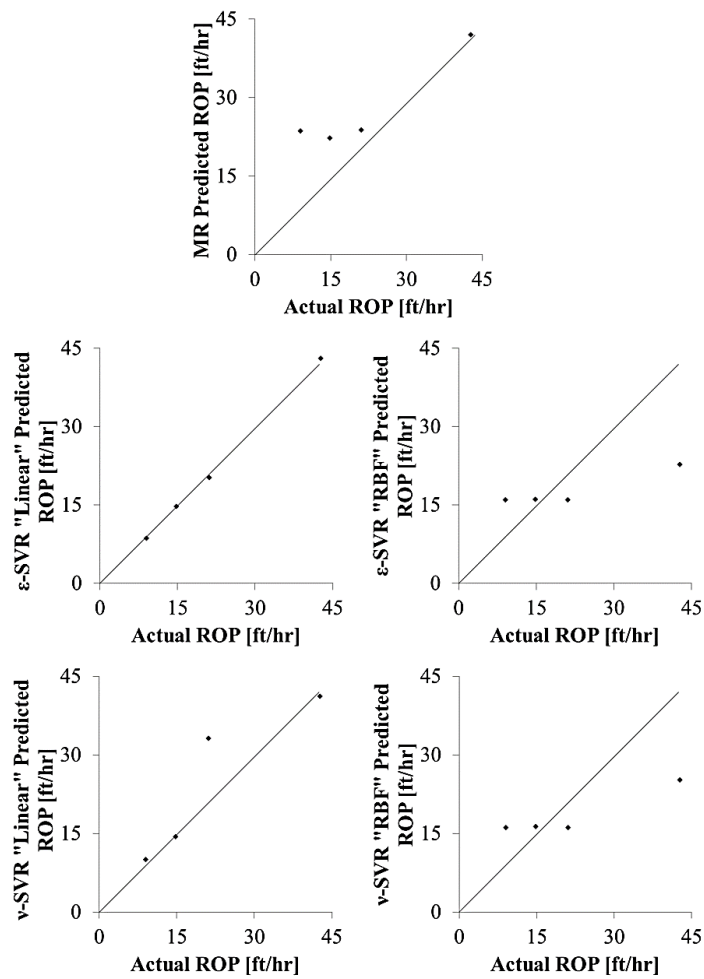


Figure 6.12 : ROP prediction comparison in Case 1 for testing #17-21-26-30.

Correspondingly to previous scenarios, ε -SVR with linear kernel is the best predictor. Likewise, ε -SVR with linear kernel has all the minimum errors and maximum Pseudo- R^2 value. In addition, the best predictors are seemed to be linear kernel methods rather than MR. Similarly, RBF kernels do not seemed to be suitable for this case as good predictive methods.

The cumulative comparative statistical results for Case 1 are placed on several charts. Negative values are excluded from the graphics. The comparison of the scenarios in Case 1 in terms of RSS_{model} is shown in Figure 6.13.

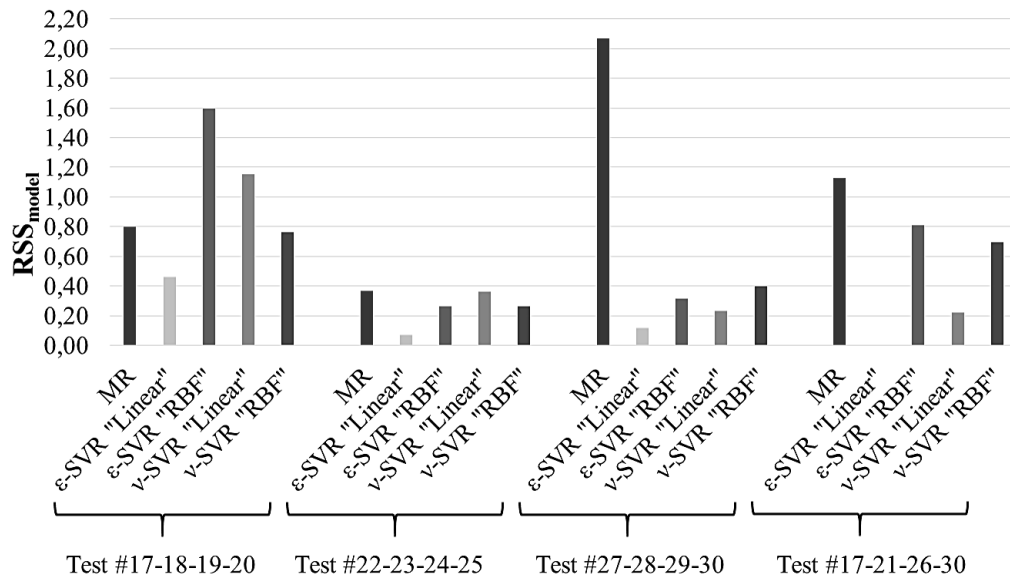


Figure 6.13 : RSS_{model} values for Case 1.

It can be seen from Figure 6.13 that MR and RBF kernels have the maximum RSS values. On the other hand, linear kernels have less RSS values, especially ε -SVR with linear kernel has the minimum RSS for all scenarios.

The comparison of the scenarios in Case 1 in terms of CV error is shown in Figure 6.14.

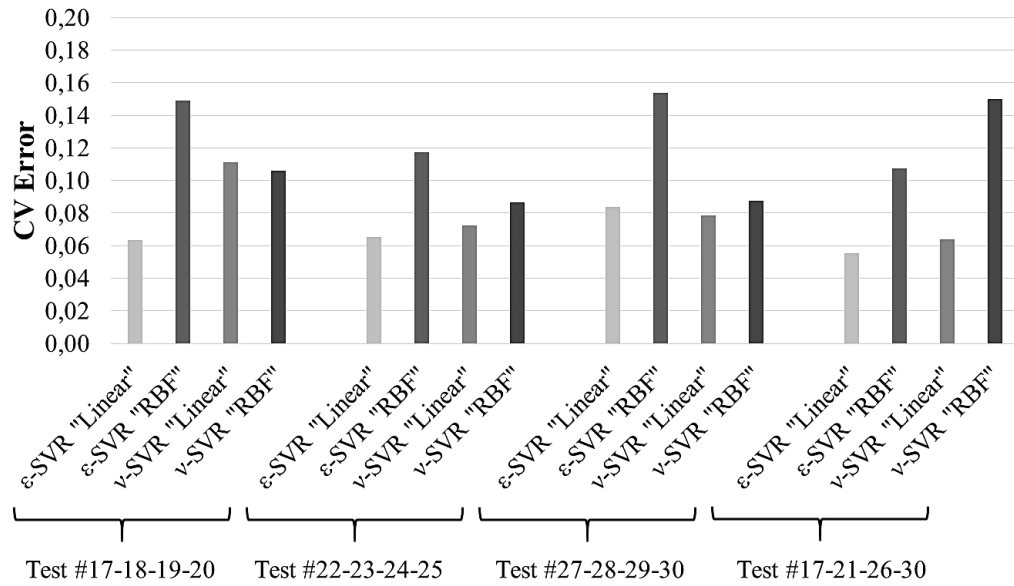


Figure 6.14 : CV error values for Case 1.

In overall, ϵ -SVR with linear kernel has the best cross validation result among other SVR methods.

The testing errors of Case 1 in terms of RMSE is given in Figure 6.15.

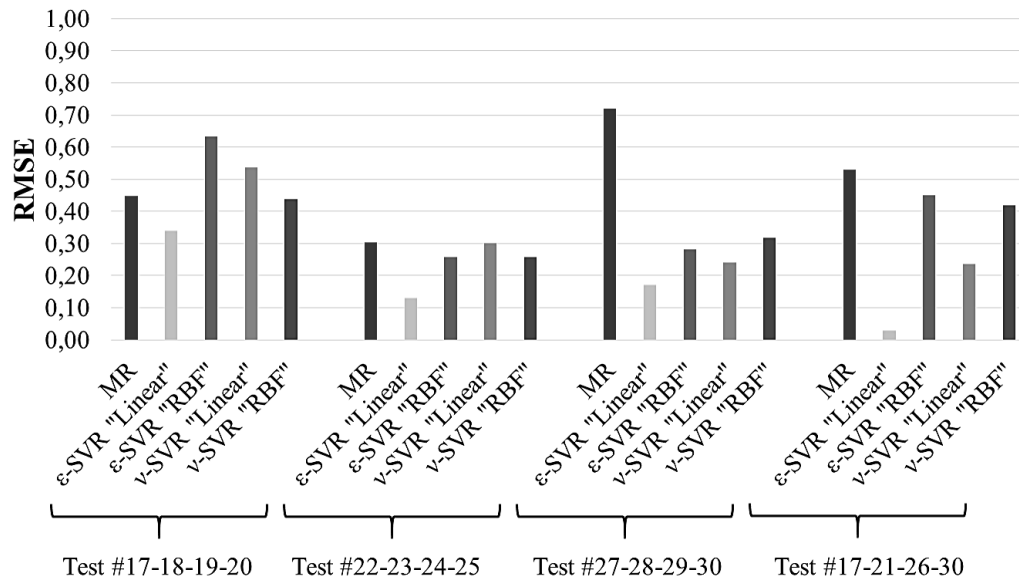


Figure 6.15 : Testing errors for Case 1.

As shown in Figure 6.15, ϵ -SVR with linear kernel has the minimum testing error.

The Pseudo- R^2 comparison for Case 1 is given in Figure 6.16.

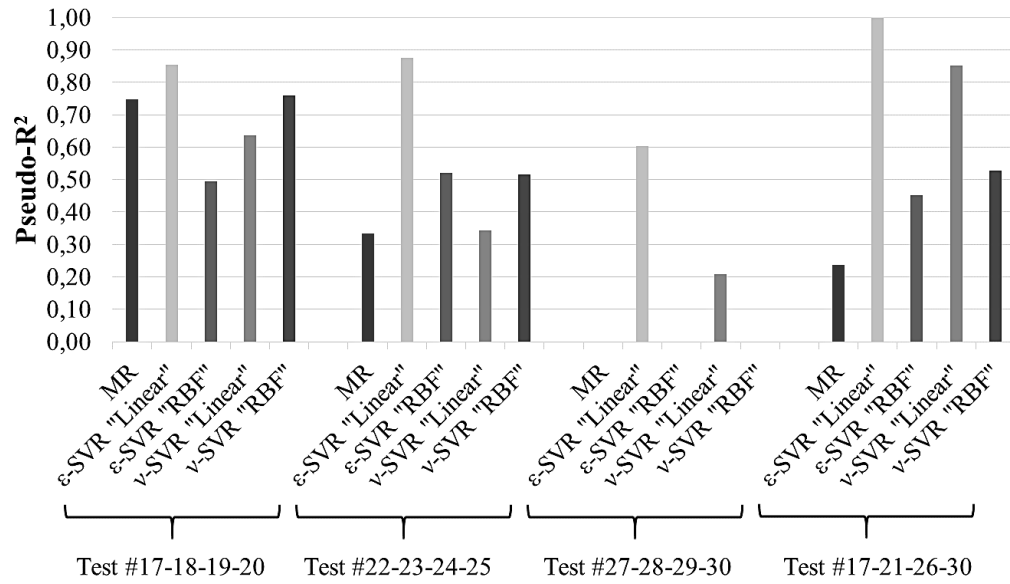


Figure 6.16 : Pseudo- R^2 values for Case 1.

It can be understood from Figure 6.16 that ϵ -SVR with linear kernel has the highest Pseudo- R^2 values for all the scenarios in Case 1. Additionally, data are shown includes negative values; i.e. the predictor model completely fails.

6.2 Case 2: Training Bottom 16 Depth Interval

In Case 2, last 16 data points of Table A.1 are used as training data set. The coefficients of MR analysis for this training data are given in Table 6.18.

Table 6.18 : Multiple regression coefficients of Case 2.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
3,176E+00	2,228E-04	3,094E-04	4,142E-05	2,972E-01	-3,543E-01	3,832E-01	2,299E-01

6.2.1 Testing #11-12-13-14

In this testing scenario, the ROP values of the inputs numbered #11-12-13-14 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.17.

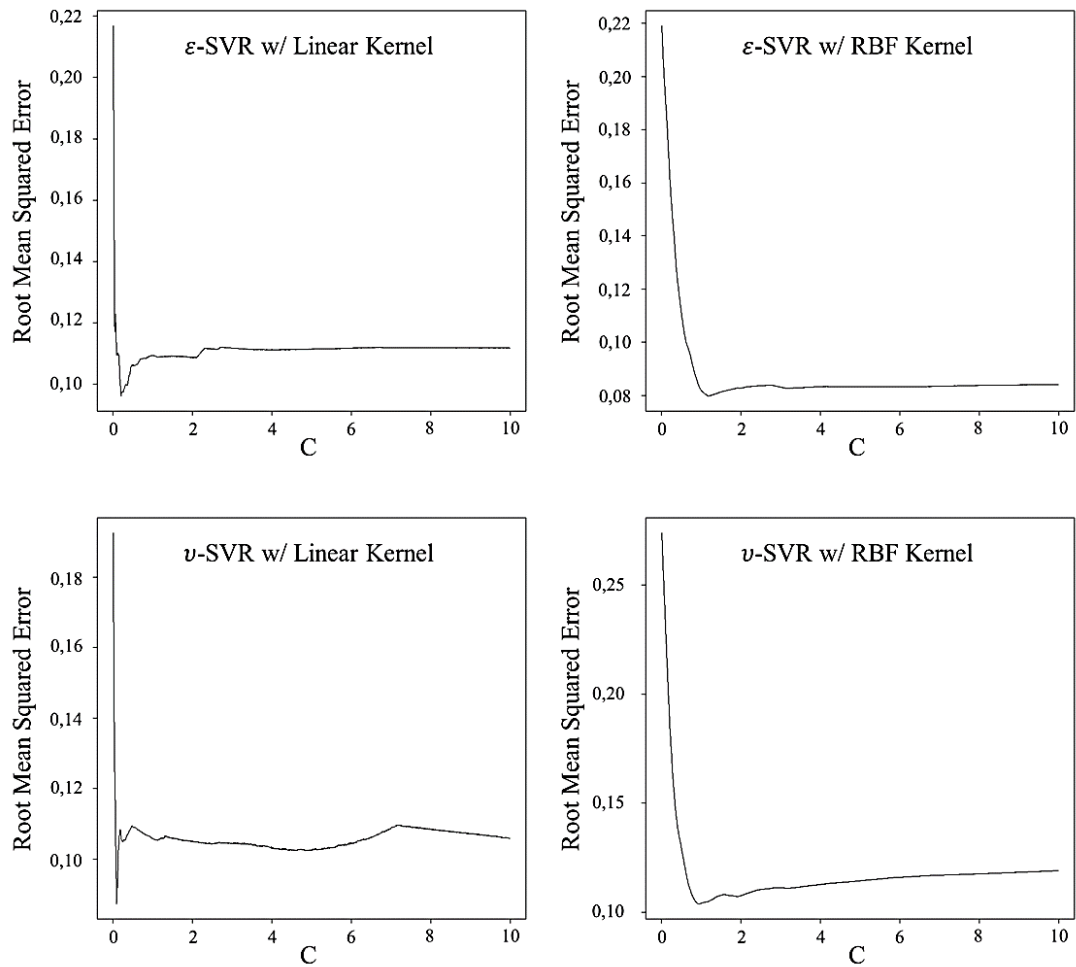


Figure 6.17 : Cost graphics of Case 2 for testing #11-12-13-14.

The basic statistical analysis of the whole data set is given in Table 6.19.

Table 6.19 : Statistics of the data set in Case 2 for testing #11-12-13-14.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,860	-4226	4656	-15129	-0,508	-0,302	-0,338	-0,286
Median	2,721	-3902	5201	-12258	-0,532	-0,274	-0,300	-0,180
Variance	0,273	4731468	3380847	68364534	0,108	0,013	0,054	0,257

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.20.

Table 6.20 : Statistical results of MR in Case 2 for learning and testing #11-12-13-14.

RSS _{default}	R ²	Adjusted R ²	p-value
2,17	0,88	0,78	0,003

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.21.

Table 6.21 : Statistical results of the methods in Case 2 for testing #11-12-13-14.

	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,42	0,77	1,26	1,82	1,36
Pseudo-R ²	0,81	0,64	0,42	0,16	0,37
C		0,205	1,17	0,09	0,94
# of SV		14	14	12	14
CV Error		0,10	0,08	0,09	0,10
RMSE	0,32	0,44	0,56	0,67	0,58

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.22. The ROP values are in unit of [ft/hr].

Table 6.22 : ROP predictions in Case 2 for #11-12-13-14.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
11	9,9	30,0	16,7	35,3	17,0	14,0
12	8,3	18,1	16,6	21,7	16,9	13,5
13	6,6	8,6	16,1	12,8	16,5	6,2
14	7,5	10,0	16,2	15,3	16,7	9,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.18.

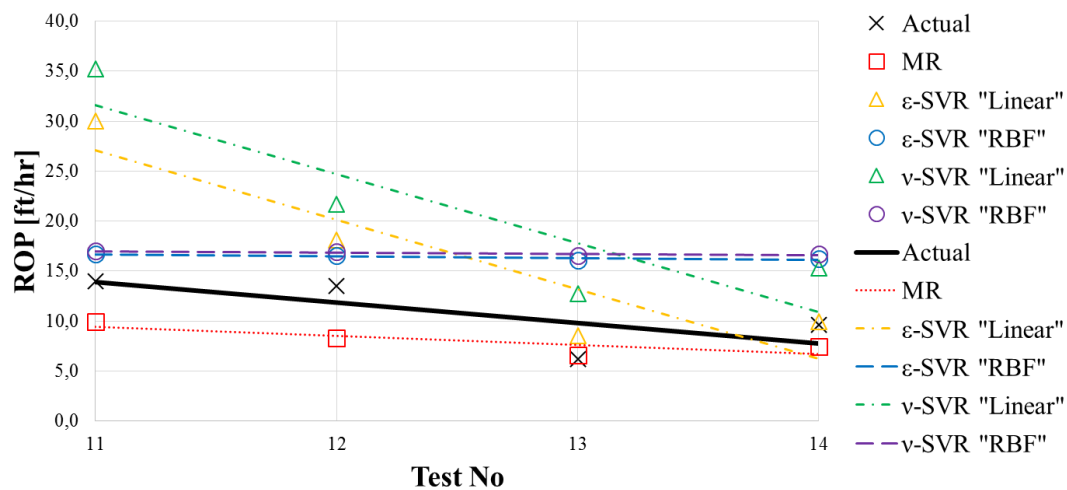


Figure 6.18 : ROP prediction trends in Case 2 for testing #11-12-13-14.

The comparison of each prediction with actual ROP is plotted in Figure 6.19.

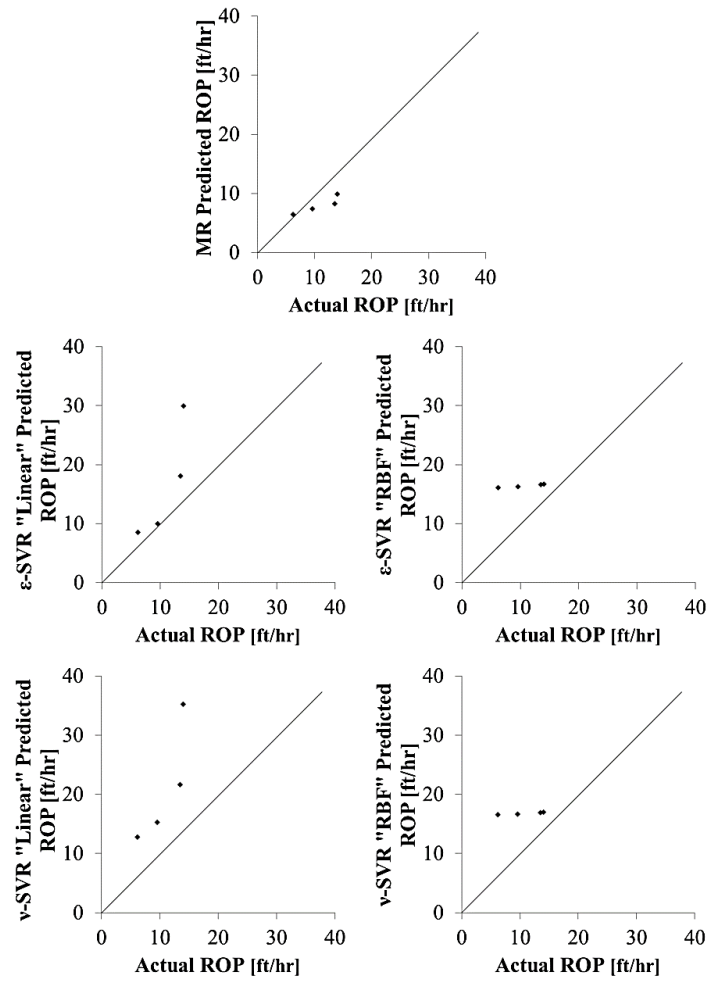


Figure 6.19 : ROP prediction comparison in Case 2 for testing #11-12-13-14.

It can be seen from Figure 6.18 and 6.19 that the best predictive method in Case 2 for testing #11-12-13-14 is MR, since it has the minimum error and maximum Pseudo- R^2 ratio among the other methods.

6.2.2 Testing #6-7-8-9

In this testing scenario, the ROP values of the inputs numbered #6-7-8-9 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.20.

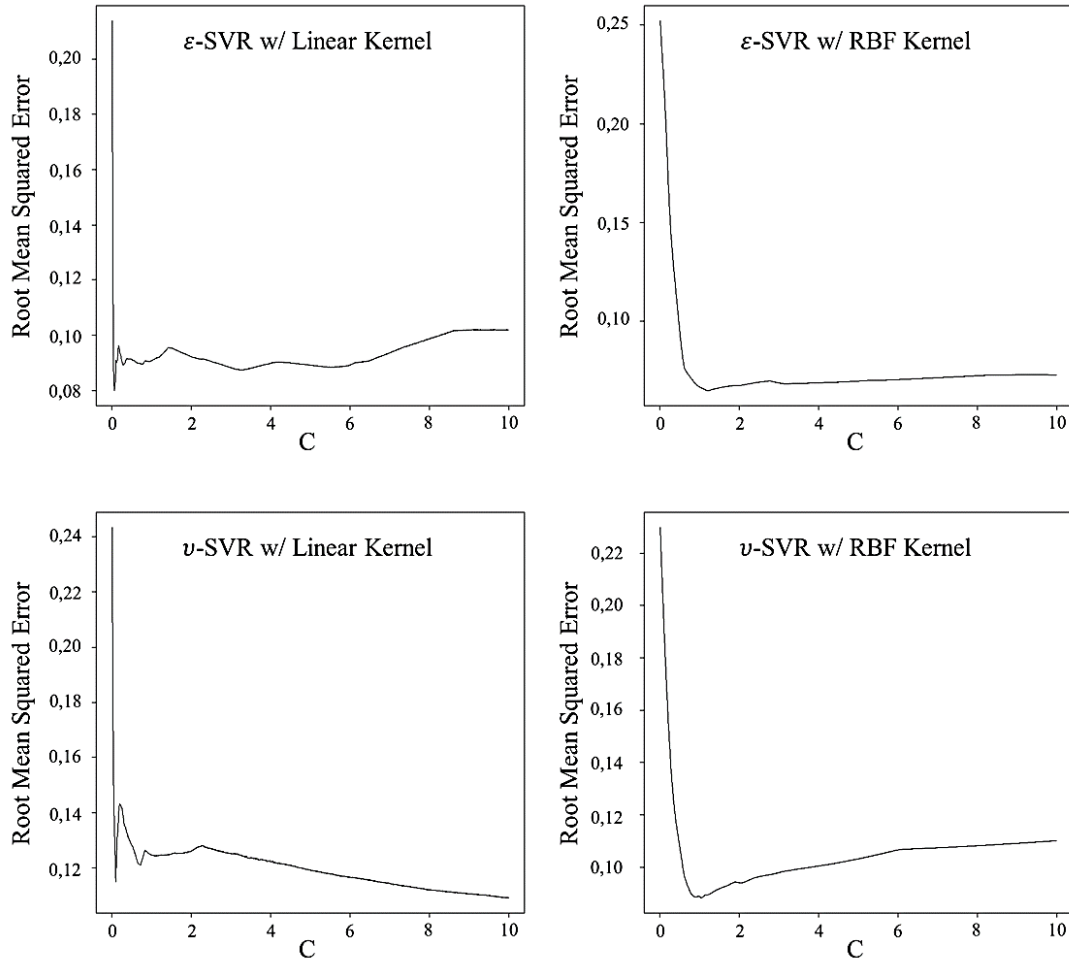


Figure 6.20 : Cost graphics of Case 2 for testing #6-7-8-9.

The basic statistical analysis of the whole data set is given in Table 6.23.

Table 6.23 : Statistics of the data set in Case 2 for testing #6-7-8-9.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,948	-4006	4394	-11378	-0,550	-0,295	-0,422	-0,282
Median	2,869	-3902	5201	-9262	-0,532	-0,274	-0,410	-0,165
Variance	0,189	6043537	5236731	34618792	0,126	0,010	0,054	0,259

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.24.

Table 6.24 : Statistical results of MR in Case 2 for learning and testing #6-7-8-9.

RSS _{default}	R ²	Adjusted R ²	p-value
0,27	0,88	0,78	0,003

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.25.

Table 6.25 : Statistical results of the methods in Case 2 for testing #6-7-8-9.

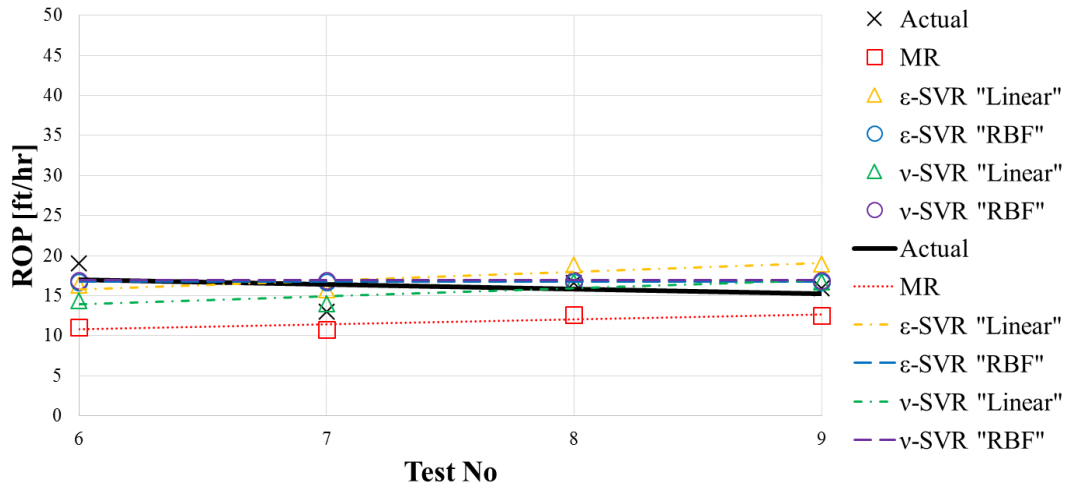
	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,46	0,10	0,08	0,09	0,09
Pseudo-R ²	-0,72	0,61	0,69	0,67	0,68
C		3,2	1,21	9,98	1,04
# of SV		14	14	13	13
CV Error		0,08	0,06	0,11	0,09
RMSE	0,34	0,16	0,14	0,15	0,15

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.26. The ROP values are in unit of [ft/hr].

Table 6.26 : ROP predictions in Case 2 for #6-7-8-9.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
6	11,1	16,3	16,7	14,3	16,9	19,0
7	10,8	15,7	16,7	14,0	16,9	13,0
8	12,6	18,8	16,7	16,7	16,9	16,6
9	12,5	18,9	16,7	16,7	16,9	15,9

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.21.

**Figure 6.21** : ROP prediction trends in Case 2 for testing #6-7-8-9.

The comparison of each prediction with actual ROP is plotted in Figure 6.22.

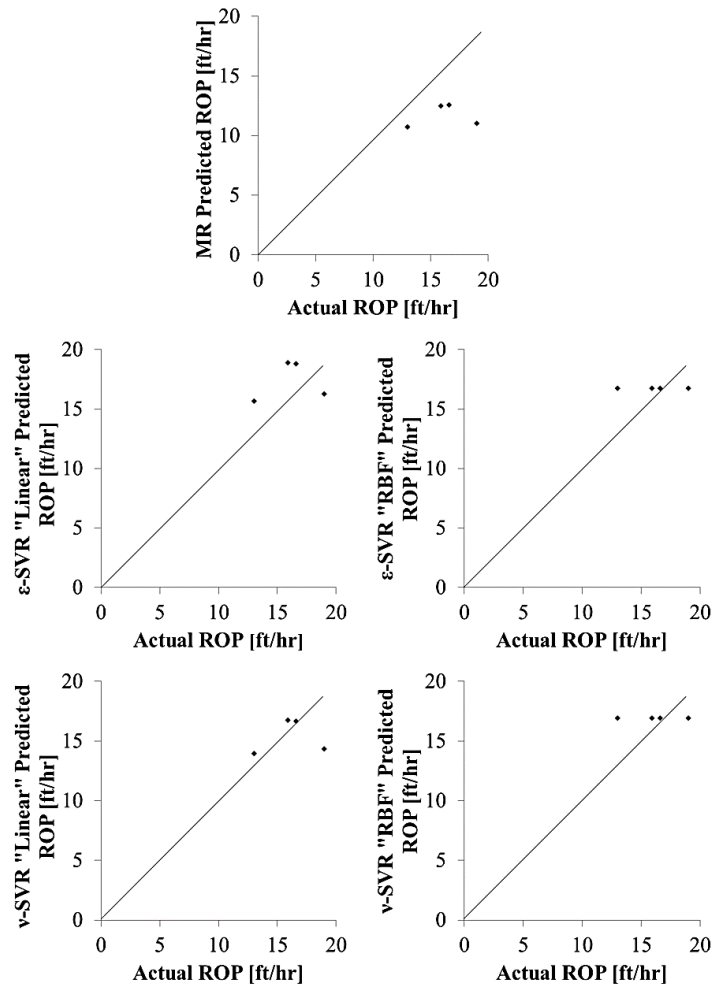


Figure 6.22 : ROP prediction comparison in Case 2 for testing #6-7-8-9.

For this scenario, it is hard to make comments on the best prediction. Maximum correlation seems to be v -SVR with linear kernel from Figure 6.22. Minimum RSS_{model} value for v -SVR with linear kernel supports this idea. However, v -SVR with linear kernel has the highest CV error. Additionally, in Figure 6.21, it can be seen that the behaviour of trends of all predictive models do not resemble to actual ROP's. For this situation, it is not quite possible to determine the best predictor.

6.2.3 Testing #1-2-3-4

In this testing scenario, the ROP values of the inputs numbered #1-2-3-4 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.23.

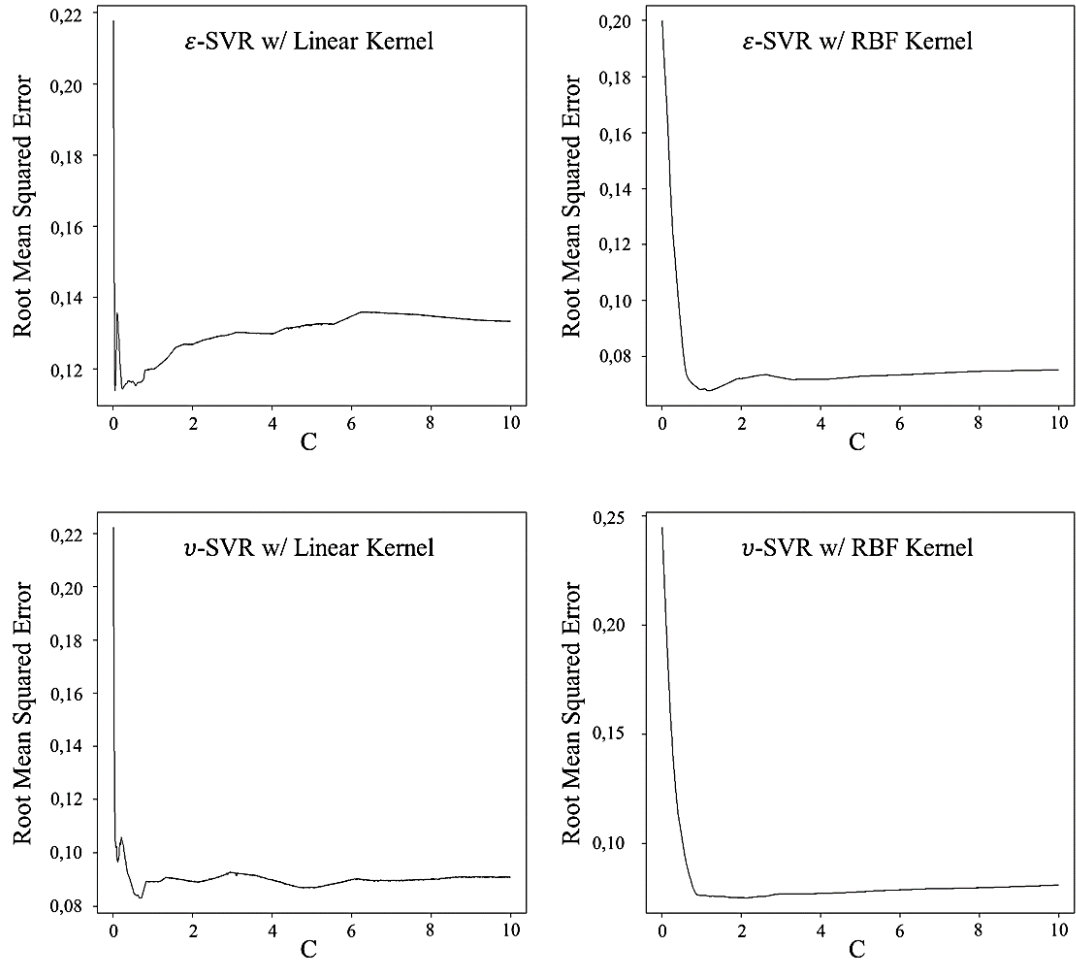


Figure 6.23 : Cost graphics of Case 2 for testing #1-2-3-4.

The basic statistical analysis of the whole data set is given in Table 6.27.

Table 6.27 : Statistics of the data set in Case 2 for testing #1-2-3-4.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,952	-3849	4363	-11455	-0,612	-0,209	-0,410	-0,285
Median	2,936	-3902	5201	-9262	-0,507	-0,255	-0,390	-0,195
Variance	0,209	7247692	5510558	33773810	0,225	0,041	0,061	0,258

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.28.

Table 6.28 : Statistical results of MR in Case 2 for learning and testing #1-2-3-4.

RSS _{default}	R ²	Adjusted R ²	p-value
0,63	0,88	0,78	0,003

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.29.

Table 6.29 : Statistical results of the methods in Case 2 for testing #1-2-3-4.

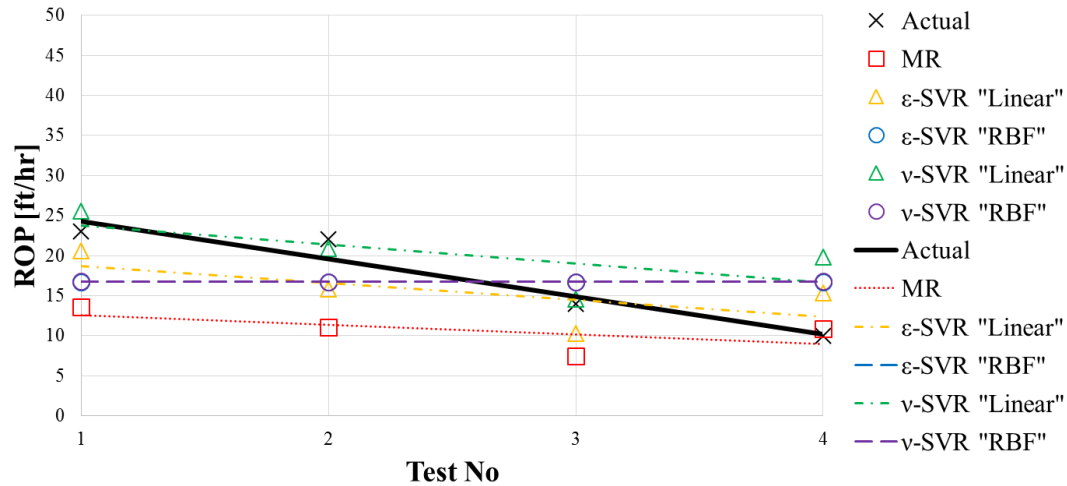
	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	1,16	0,40	0,47	0,48	0,47
Pseudo-R ²	-0,84	0,36	0,25	0,23	0,25
C		2	1,185	2	2,12
# of SV		13	13	13	14
CV Error		0,11	0,07	0,08	0,09
RMSE	0,54	0,32	0,34	0,35	0,34

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.30. The ROP values are in unit of [ft/hr].

Table 6.30 : ROP predictions in Case 2 for #1-2-3-4.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
1	13,6	20,6	16,7	25,6	16,7	23,0
2	11,1	15,9	16,7	20,9	16,7	22,0
3	7,4	10,2	16,7	14,6	16,7	14,0
4	10,8	15,4	16,7	19,8	16,7	10,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.24.

**Figure 6.24** : ROP prediction trends in Case 2 for testing #1-2-3-4.

The comparison of each prediction with actual ROP is plotted in Figure 6.25.

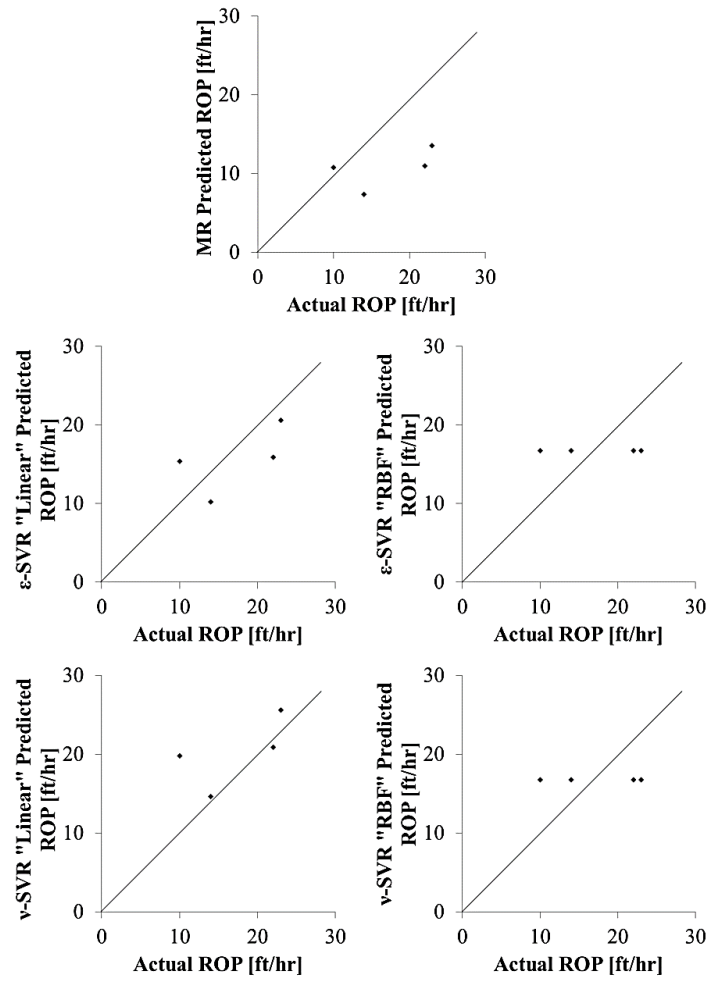


Figure 6.25 : ROP prediction comparison in Case 2 for testing #1-2-3-4.

It can be understood from Figure 6.25 that, ν -SVR with linear kernel has the best predictions, as also can be seen from Figure 6.24. Furthermore, the minimum errors and maximum Pseudo- R^2 value are in ν -SVR with linear kernel. Additionally, the best predictors are seemed to be linear kernel methods as in the third scenario. It is quite obvious that RBF kernels do not generate meaningful results.

6.2.4 Testing #1-5-10-14

In this testing scenario, the ROP values of the inputs numbered #1-5-10-14 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.26.

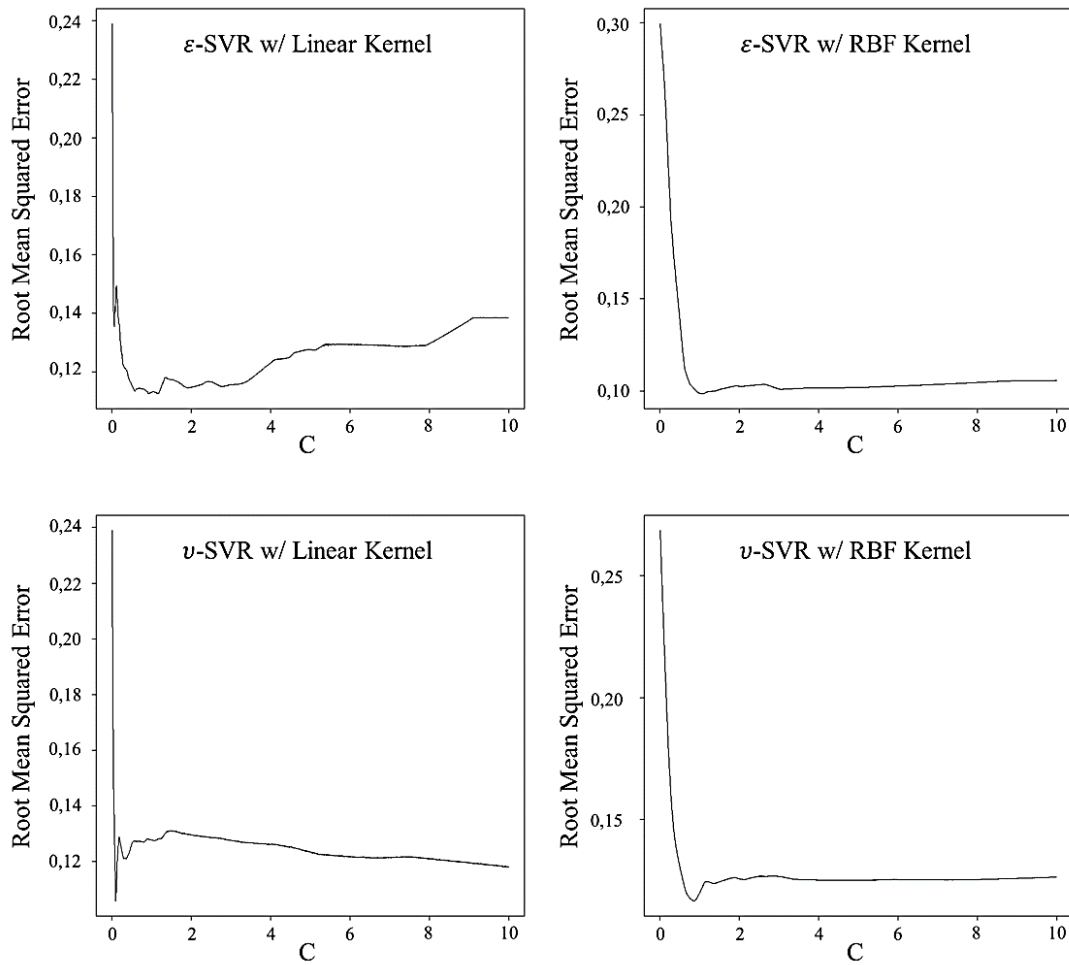


Figure 6.26 : Cost graphics of Case 2 for testing #1-5-10-14.

The basic statistical analysis of the whole data set is given in Table 6.31.

Table 6.31 : Statistics of the data set in Case 2 for testing #1-5-10-14.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,940	-4014	4510	-12705	-0,533	-0,266	-0,398	-0,280
Median	2,851	-3902	5201	-10219	-0,507	-0,268	-0,365	-0,165
Variance	0,209	6164838	4509132	55304663	0,141	0,017	0,064	0,260

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.32.

Table 6.32 : Statistical results of MR in Case 2 for learning and testing #1-5-10-14.

RSS _{default}	R ²	Adjusted R ²	p-value
0,66	0,88	0,78	0,003

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.33.

Table 6.33 : Statistical results of the methods in Case 2 for testing #1-5-10-14.

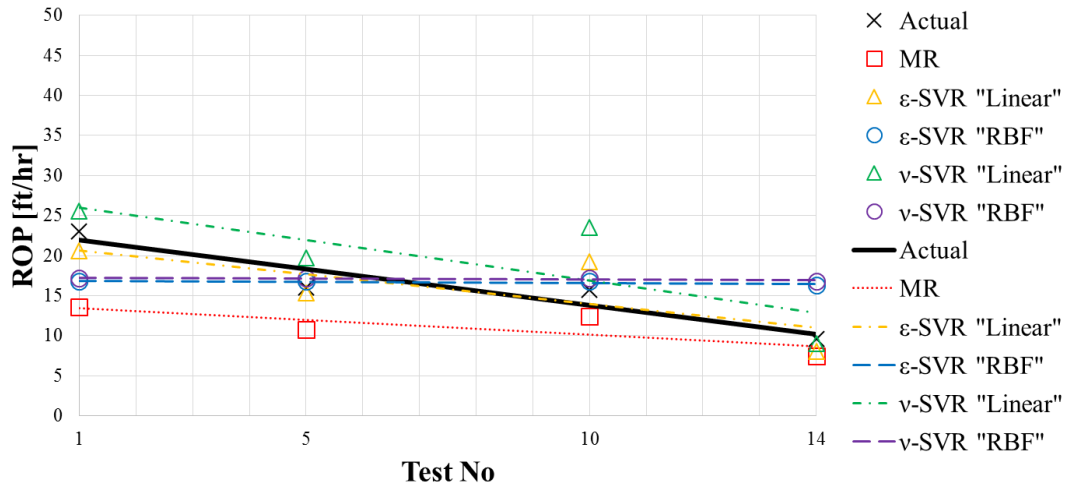
	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,55	0,09	0,39	0,22	0,41
Pseudo-R ²	0,16	0,87	0,41	0,66	0,37
C		2	1,05	2	0,845
# of SV		13	14	13	14
CV Error		0,11	0,10	0,11	0,12
RMSE	0,37	0,15	0,31	0,24	0,32

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.34. The ROP values are in unit of [ft/hr].

Table 6.34 : ROP predictions in Case 2 for #1-5-10-14.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
1	13,6	20,6	16,8	25,6	17,2	23,0
5	10,7	15,4	16,8	19,7	17,2	16,0
10	12,5	19,2	16,8	23,5	17,2	15,7
14	7,5	8,0	16,3	9,0	16,8	9,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.27.

**Figure 6.27 :** ROP prediction trends in Case 2 for testing #1-5-10-14.

The comparison of each prediction with actual ROP is plotted in Figure 6.28.

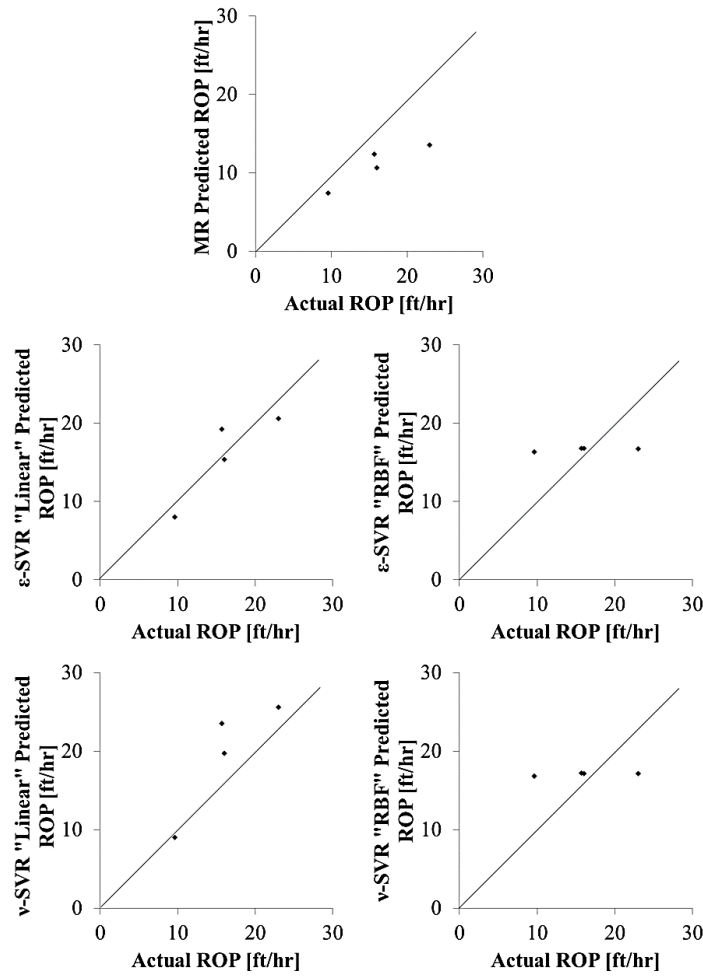


Figure 6.28 : ROP prediction comparison in Case 2 for testing #1-5-10-14.

For testing #1-5-10-14, ε -SVR with linear kernel is seemed to be the best predictor. Likewise, ε -SVR with linear kernel has all the minimum errors and maximum Pseudo- R^2 value. In addition, the best predictors are seemed to be linear kernel methods rather than MR. Similarly, RBF kernels do not seemed to be suitable for this case as good predictive methods.

The cumulative comparative statistical results for Case 2 are placed on several charts. Negative values are excluded from the graphics. The comparison of the scenarios in Case 2 in terms of RSS_{model} is shown in Figure 6.29.

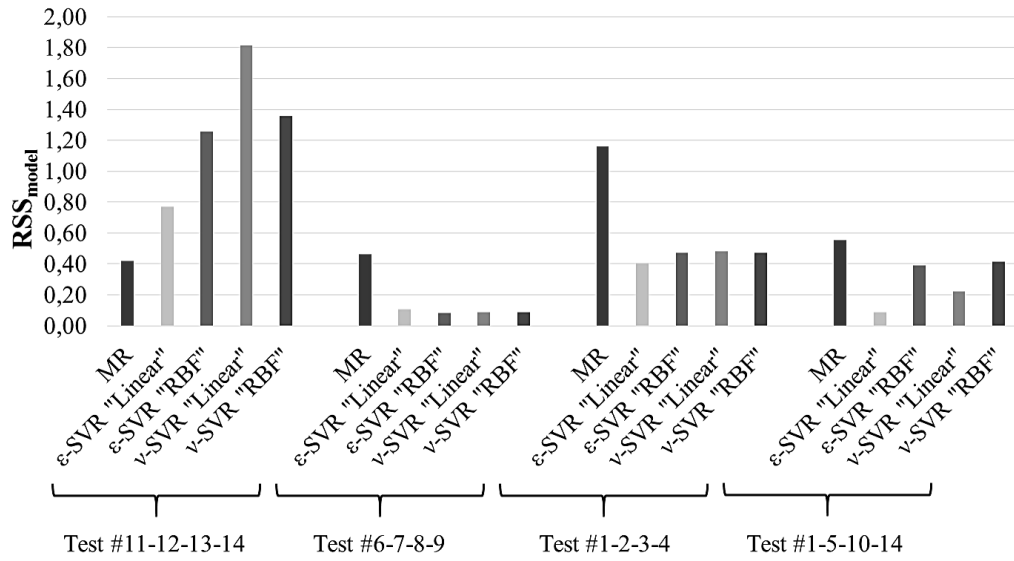


Figure 6.29 : RSS_{model} values for Case 2.

It can be seen from Figure 6.29 that MR has the maximum RSS values in most of the scenarios. On the other hand, linear kernels have less RSS values, especially ϵ -SVR with linear kernel has the minimum RSS for the most scenarios.

The comparison of the scenarios in Case 2 in terms of CV error is shown in Figure 6.30.

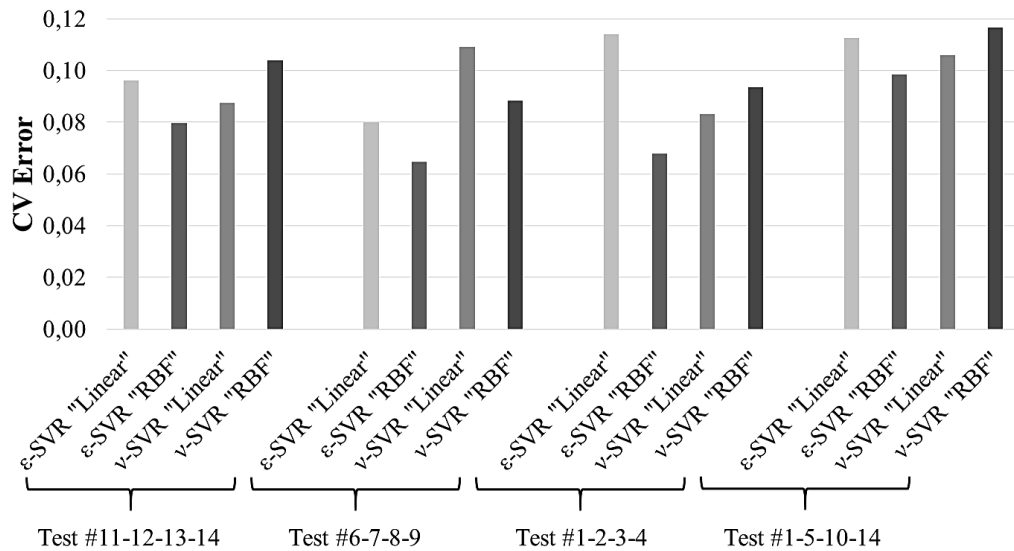


Figure 6.30 : CV error values for Case 2.

In overall, ϵ -SVR with RBF kernel has the best cross validation result among other SVR methods.

The testing errors of Case 2 in terms of RMSE is given in Figure 6.31.

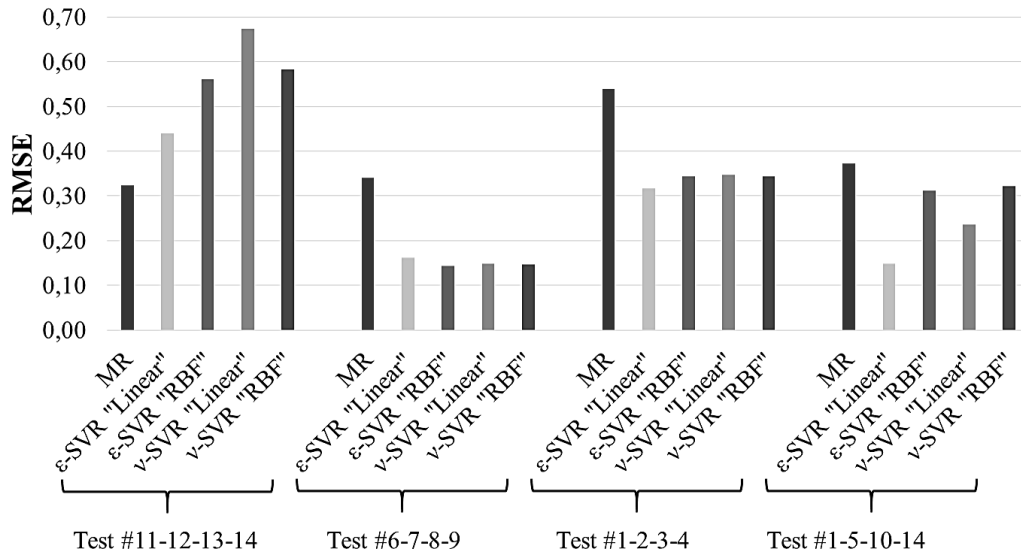


Figure 6.31 : Testing errors for Case 2.

As shown in Figure 6.31, ϵ -SVR with linear kernel has the minimum testing error for the most of the scenarios.

The Pseudo- R^2 comparison for Case 2 is given in Figure 6.32.

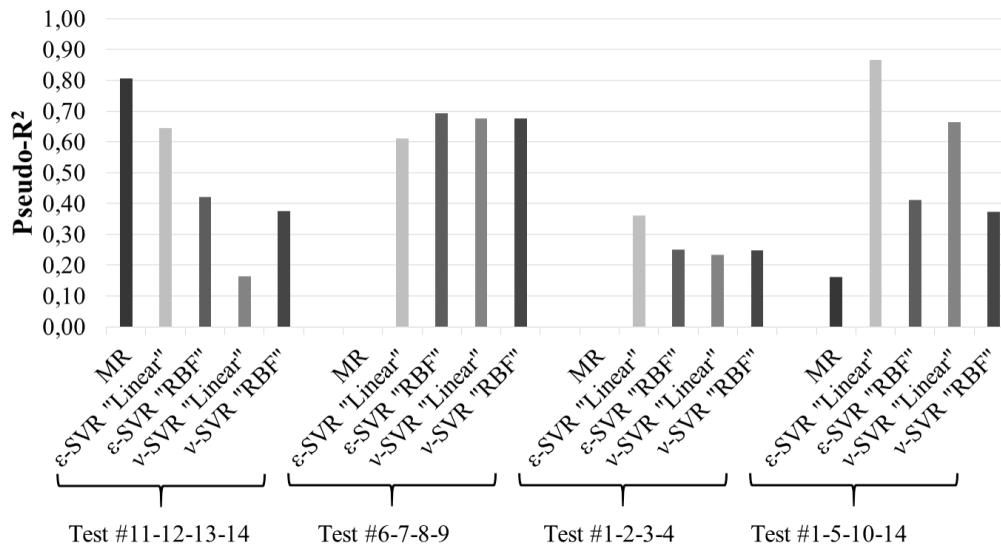


Figure 6.32 : Pseudo- R^2 values for Case 2.

It can be understood from Figure 6.32 that ϵ -SVR with linear kernel can be preferred as a predicting method for these kind of scenarios. However, it is seen that the results are changeable according to the tested data inputs.

6.3 Case 3: Training Odd Numbered Data Points

In Case 3, the odd numbered data points of Table A.1 are used as training data set. The coefficients of MR analysis for this training data are given in Table 6.35.

Table 6.35 : Multiple regression coefficients of Case 3.

a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
3,611E+00	2,299E-04	2,763E-04	4,796E-05	2,935E-01	9,691E-02	3,088E-01	7,415E-02

6.3.1 Testing #2-4-6-8

In this testing scenario, the ROP values of the inputs numbered #2-4-6-8 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.33.

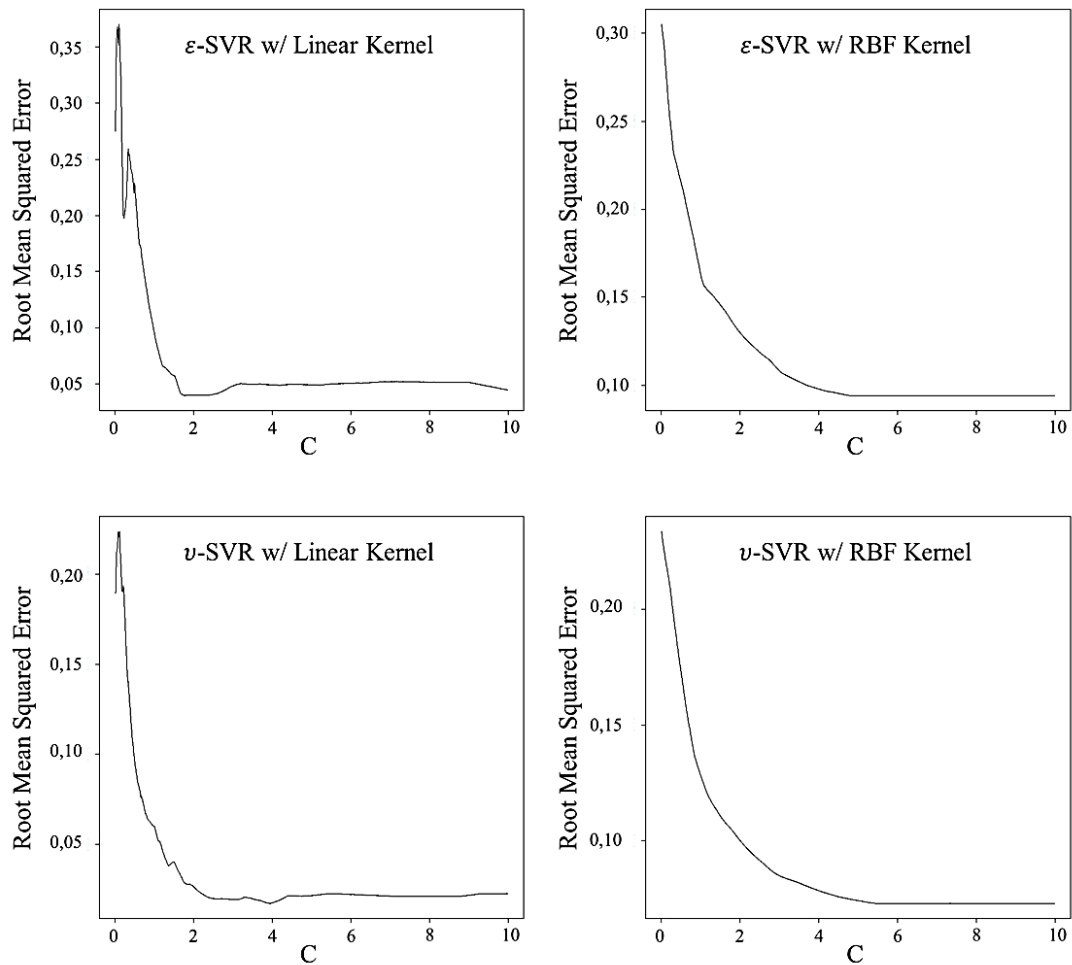


Figure 6.33 : Cost graphics of Case 3 for testing #2-4-6-8.

The basic statistical analysis of the whole data set is given in Table 6.36.

Table 6.36 : Statistics of the data set in Case 3 for testing #2-4-6-8.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,852	-2222	2390	-10711	-0,787	-0,210	-0,423	-0,126
Median	2,773	-1475	316	-7675	-0,782	-0,261	-0,380	-0,037
Variance	0,209	5227401	6941166	65892589	0,206	0,041	0,056	0,155

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.37.

Table 6.37 : Statistical results of MR in Case 3 for learning and testing #2-4-6-8.

RSS _{default}	R ²	Adjusted R ²	p-value
0,38	0,97	0,94	6,478E-05

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.38.

Table 6.38 : Statistical results of the methods in Case 3 for testing #2-4-6-8.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,24	0,27	0,37	0,24	0,43
Pseudo-R ²	0,37	0,29	0,02	0,36	-0,14
C		1,775	4,775	3,935	6,15
# of SV		12	14	13	15
CV Error		0,04	0,09	0,02	0,07
RMSE	0,24	0,26	0,30	0,25	0,33

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.39. The ROP values are in unit of [ft/hr].

Table 6.39 : ROP predictions in Case 3 for #2-4-6-8.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
2	19,1	19,2	16,8	19,0	16,6	22,0
4	15,3	16,0	16,0	15,3	16,7	10,0
6	15,6	15,8	14,6	15,7	14,1	19,0
8	16,2	16,8	15,3	16,4	16,0	16,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.34.

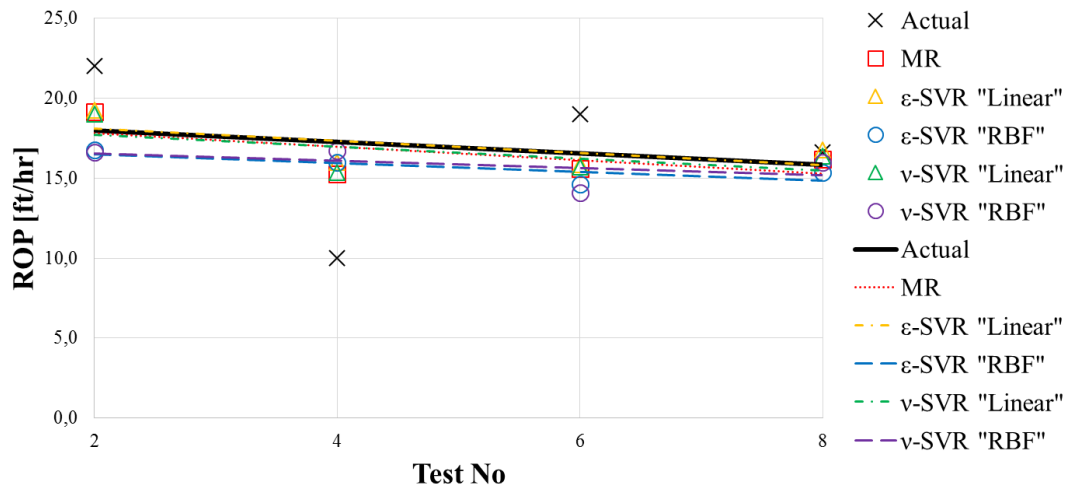


Figure 6.34 : ROP prediction trends in Case 3 for testing #2-4-6-8.

The comparison of each prediction with actual ROP is plotted in Figure 6.35.

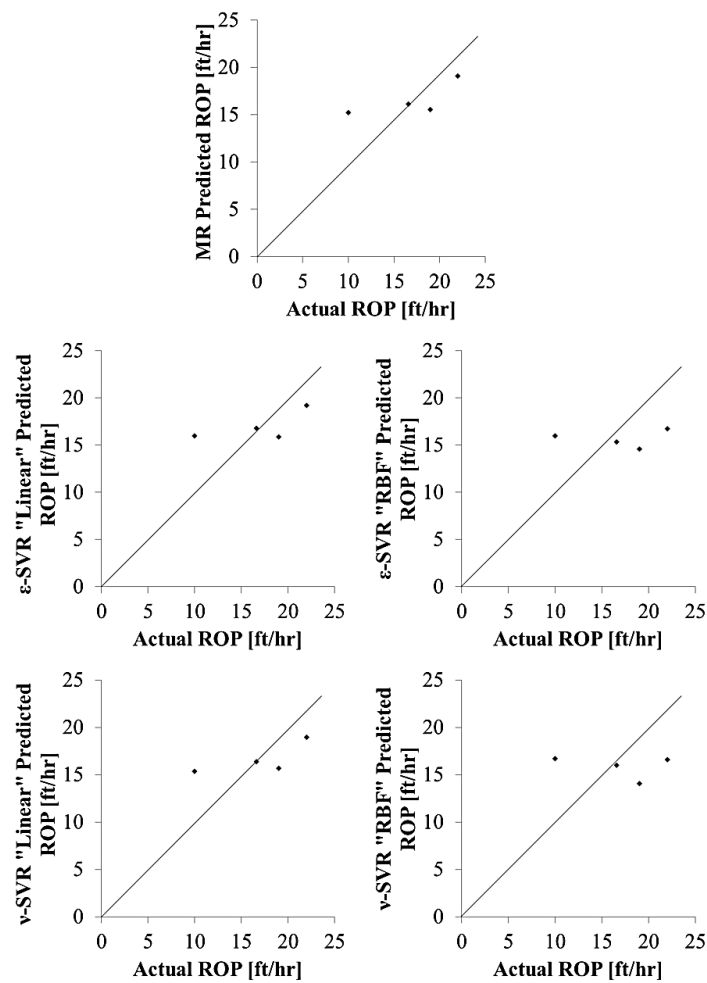


Figure 6.35 : ROP prediction comparison in Case 3 for testing #2-4-6-8.

It can be seen from Figure 6.35 and 6.34 that the best predictive methods in Case 3 for testing #2-4-6-8 are MR and SVR methods with linear kernel, especially ν -SVR, since they have the minimum error and maximum Pseudo- R^2 ratio among the other methods.

6.3.2 Testing #12-14-16-18

In this testing scenario, the ROP values of the inputs numbered #12-14-16-18 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.36.

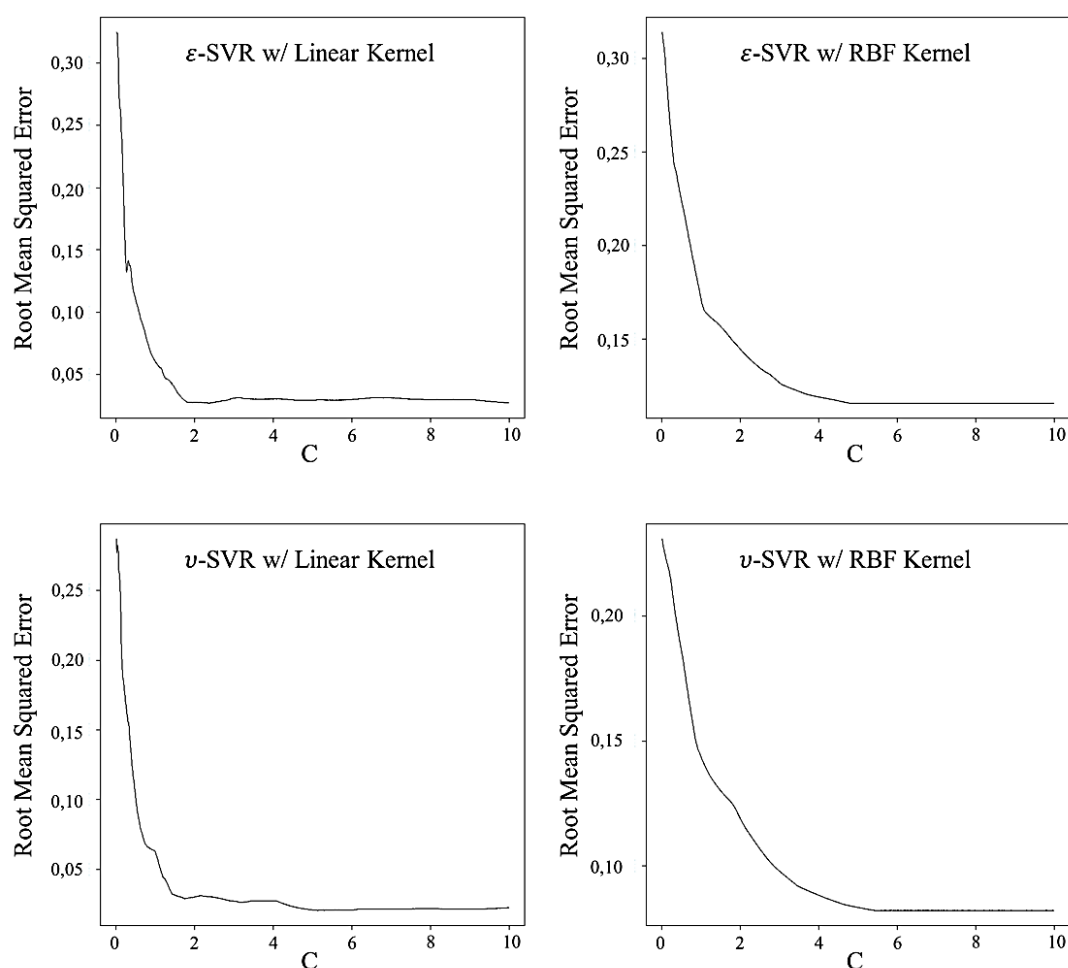


Figure 6.36 : Cost graphics of Case 3 for testing #12-14-16-18.

The basic statistical analysis of the whole data set is given in Table 6.40.

Table 6.40 : Statistics of the data set in Case 3 for testing #12-14-16-18.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,895	-2663	3064	-13182	-0,646	-0,244	-0,391	-0,112
Median	2,766	-2315	3588	-8406	-0,571	-0,261	-0,310	-0,024
Variance	0,265	4293926	6092889	86121025	0,183	0,034	0,062	0,162

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.41.

Table 6.41 : Statistical results of MR in Case 3 for testing #12-14-16-18.

RSS _{default}	R ²	Adjusted R ²	p-value
0,38	0,97	0,94	6,478E-05

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.42.

Table 6.42 : Statistical results of the methods in Case 3 for testing #12-14-16-18.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,24	0,16	0,14	0,19	0,21
Pseudo-R ²	0,83	0,88	0,90	0,86	0,85
C		2,35	4,775	5,13	5,425
# of SV		12	14	13	15
CV Error		0,03	0,12	0,02	0,08
RMSE	0,24	0,20	0,19	0,22	0,23

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.43. The ROP values are in unit of [ft/hr].

Table 6.43 : ROP predictions in Case 3 for #12-14-16-18.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
12	8,8	9,4	10,3	9,1	10,0	13,5
14	7,9	8,0	8,1	7,9	7,6	9,6
16	33,8	31,6	37,4	32,3	38,3	31,4
18	42,3	39,4	43,2	40,3	45,9	38,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.37.

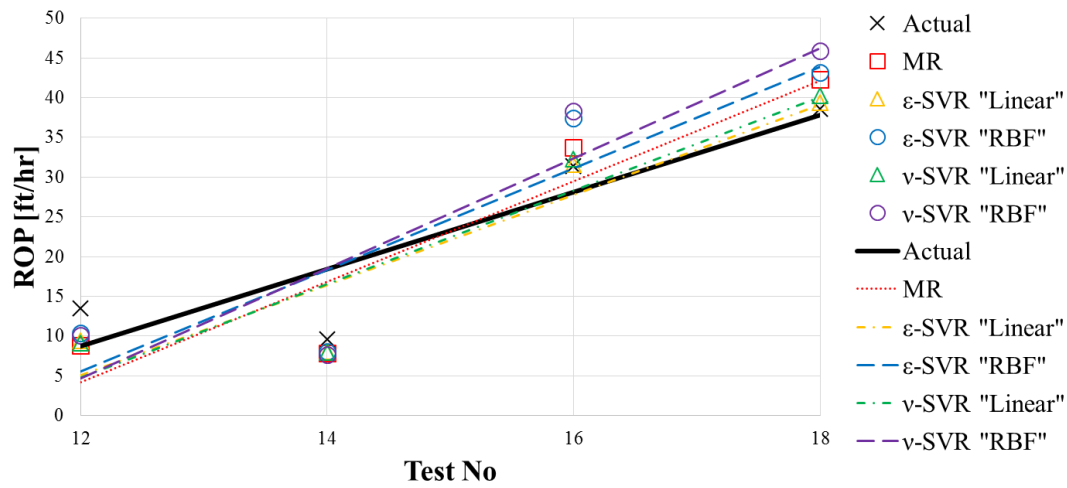


Figure 6.37 : ROP prediction trends in Case 3 for testing #12-14-16-18.

The comparison of each prediction with actual ROP is plotted in Figure 6.38.

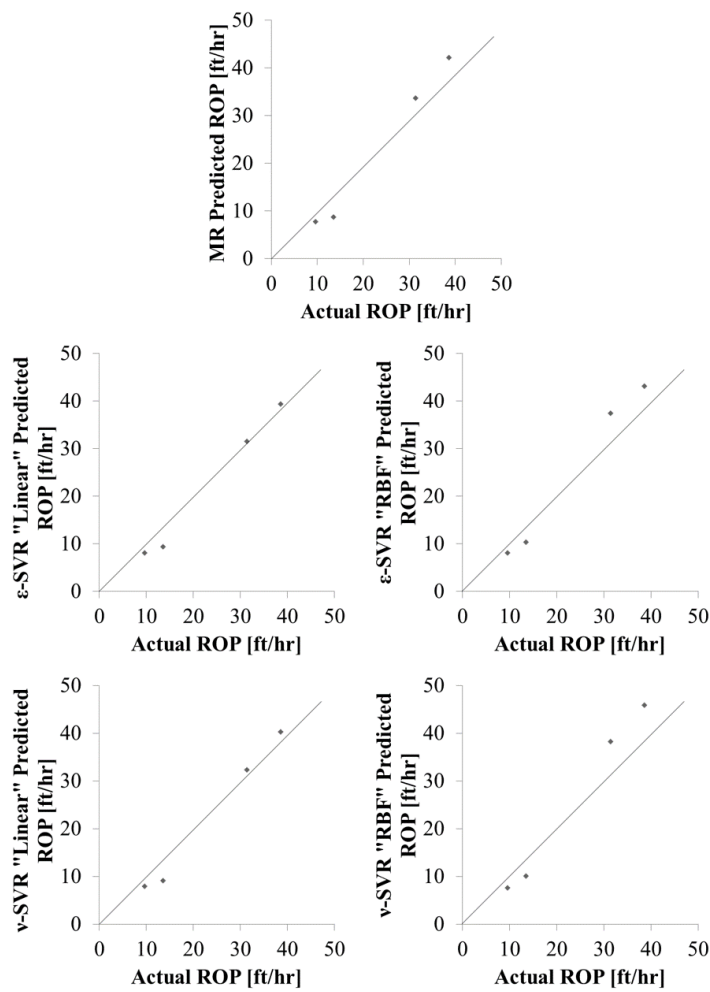


Figure 6.38 : ROP prediction comparison in Case 3 for testing #12-14-16-18.

For this scenario, the maximum correlation seems to be ε -SVR with linear kernel from Figure 6.37 and 6.38. However, ε -SVR with RBF kernel has the maximum Pseudo- R^2 value, the minimum errors, and the minimum RSS_{model} . On the other hand, residuals of ε -SVR with RBF kernel are greater than ε -SVR with linear kernel's. Thus, any judgement should lead a misleading decision for this situation. Further calculations for similar scenarios should be repeated.

6.3.3 Testing #22-24-26-28

In this testing scenario, the ROP values of the inputs numbered #22-24-26-28 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.39.

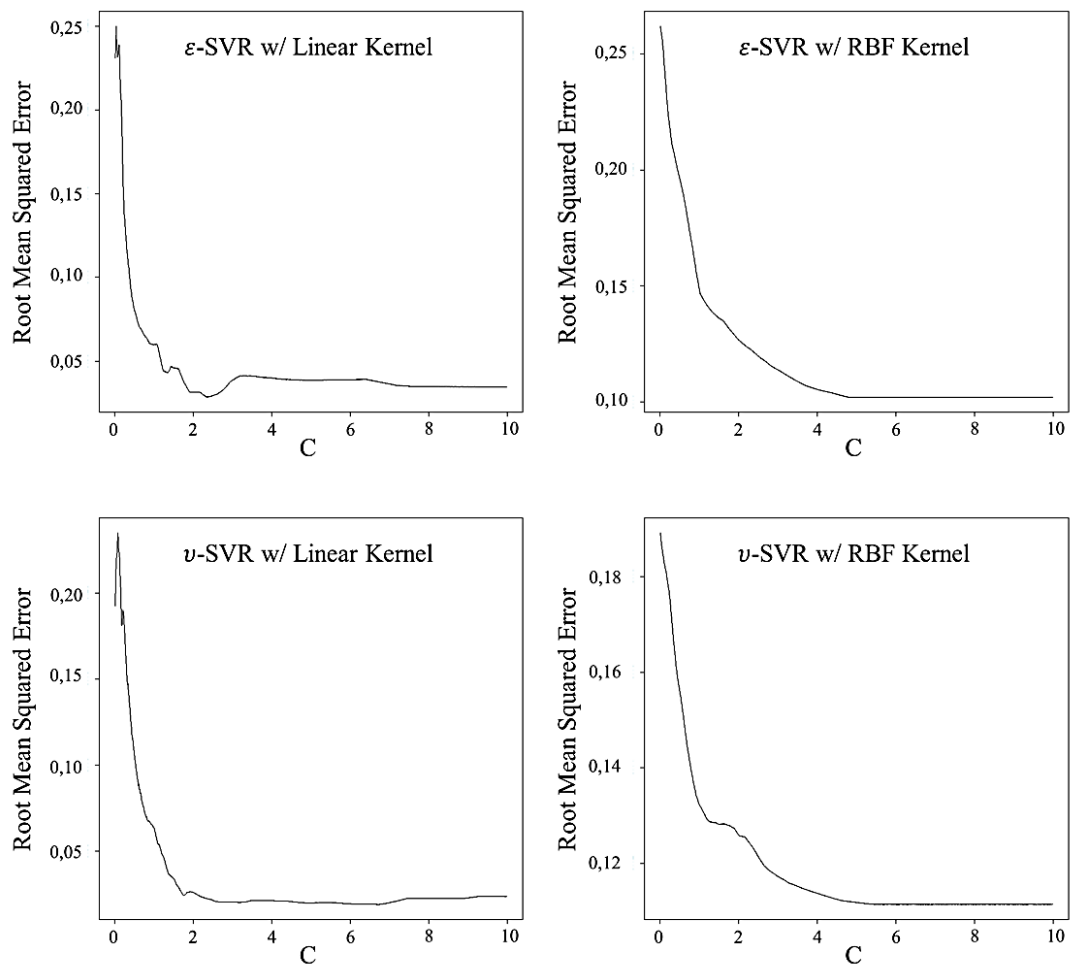


Figure 6.39 : Cost graphics of Case 3 for testing #22-24-26-28.

The basic statistical analysis of the whole data set is given in Table 6.44.

Table 6.44 : Statistics of the data set in Case 3 for testing #22-24-26-28.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,863	-3275	3587	-12054	-0,656	-0,245	-0,438	-0,225
Median	2,766	-3250	5031	-9180	-0,571	-0,261	-0,400	-0,130
Variance	0,193	5552579	6745246	59969496	0,188	0,031	0,063	0,190

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.45.

Table 6.45 : Statistical results of MR in Case 3 for learning and testing #22-24-26-28.

RSS _{default}	R ²	Adjusted R ²	p-value
0,08	0,97	0,94	6,748E-05

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.46.

Table 6.46 : Statistical results of the methods in Case 3 for testing #22-24-26-28.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,16	0,15	0,12	0,16	0,10
Pseudo-R ²	-0,95	-0,85	-0,48	-0,94	-0,25
C		2,355	4,775	6,72	6,85
# of SV		12	14	13	15
CV Error		0,11	0,10	0,02	0,11
RMSE	0,20	0,20	0,18	0,20	0,16

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.47. The ROP values are in unit of [ft/hr].

Table 6.47 : ROP predictions in Case 3 for #22-24-26-28.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
22	25,1	25,0	18,3	24,7	18,0	19,0
24	22,4	21,9	17,5	22,2	17,3	20,2
26	17,7	17,6	18,9	17,8	18,2	14,8
28	18,3	18,3	18,3	18,7	17,9	14,9

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.40.

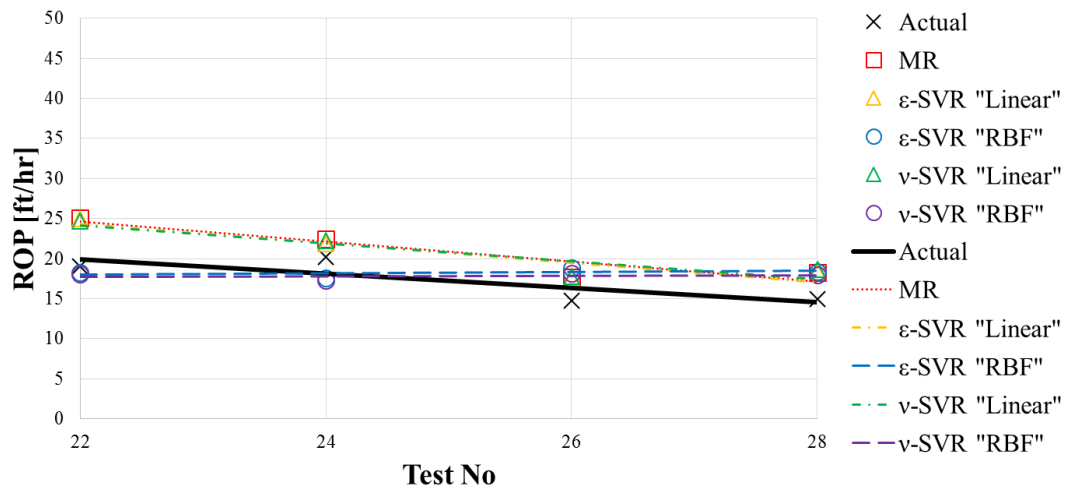


Figure 6.40 : ROP prediction trends in Case 3 for testing #22-24-26-28.

The comparison of each prediction with actual ROP is plotted in Figure 6.41.

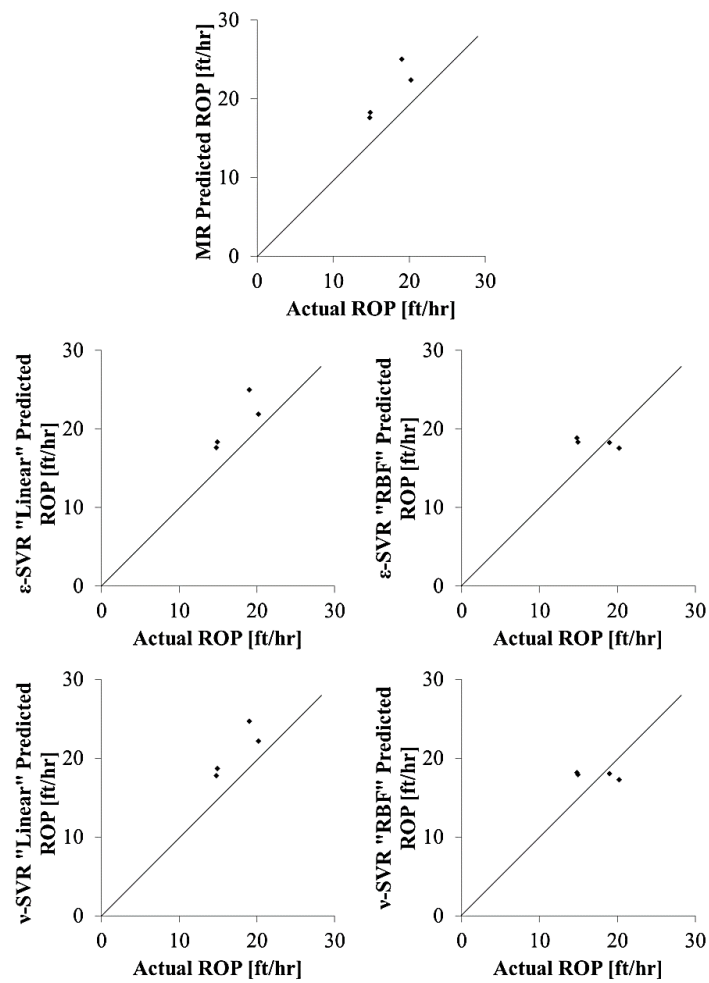


Figure 6.41 : ROP prediction comparison in Case 3 for testing #22-24-26-28.

It can be understood from Figure 6.41 that, SVR methods with RBF kernel seemed to be the best predictors. However, in Figure 6.40 it can be seen that none of the methods have good approximation to the prediction model. Furthermore, all the models have negative Pseudo- R^2 value, which means the predictor model is worse than our default model.

6.3.4 Testing #2-10-20-30

In this testing scenario, the ROP values of the inputs numbered #2-10-20-30 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.42.

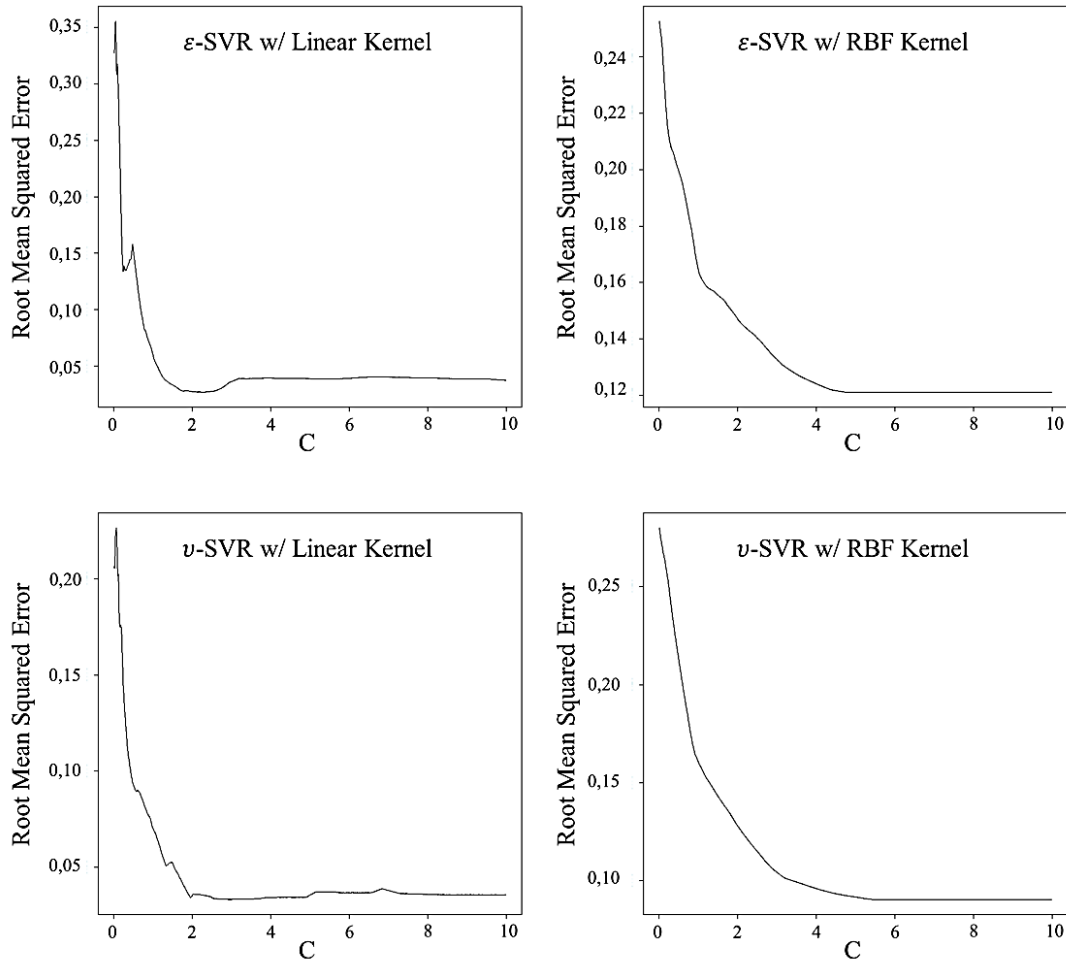


Figure 6.42 : Cost graphics of Case 3 for testing #2-10-20-30.

The basic statistical analysis of the whole data set is given in Table 6.48.

Table 6.48 : Statistics of the data set in Case 3 for testing #2-10-20-30.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,822	-2934	3040	-11500	-0,711	-0,228	-0,431	-0,231
Median	2,754	-2315	3588	-7950	-0,678	-0,261	-0,400	-0,120
Variance	0,221	8011840	7518806	72720305	0,199	0,045	0,065	0,235

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.49.

Table 6.49 : Statistical results of MR in Case 3 for learning and testing #2-10-20-30.

RSS _{default}	R ²	Adjusted R ²	p-value
0,63	0,97	0,94	6,478E-05

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.50.

Table 6.50 : Statistical results of the methods in Case 3 for testing #2-10-20-30.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,53	0,41	0,49	0,46	0,51
Pseudo-R ²	0,15	0,35	0,23	0,28	0,19
C		2,235	4,725	2,99	5,505
# of SV		12	14	13	15
CV Error		0,03	0,12	0,03	0,09
RMSE	0,37	0,32	0,35	0,34	0,36

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.51. The ROP values are in unit of [ft/hr].

Table 6.51 : ROP predictions in Case 3 for #2-10-20-30.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
2	19,1	19,2	16,8	19,0	16,6	22,0
10	16,9	17,1	15,2	17,1	15,6	15,7
20	24,8	23,2	18,8	23,9	19,2	12,5
30	7,4	8,7	14,8	8,3	14,8	9,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.43.

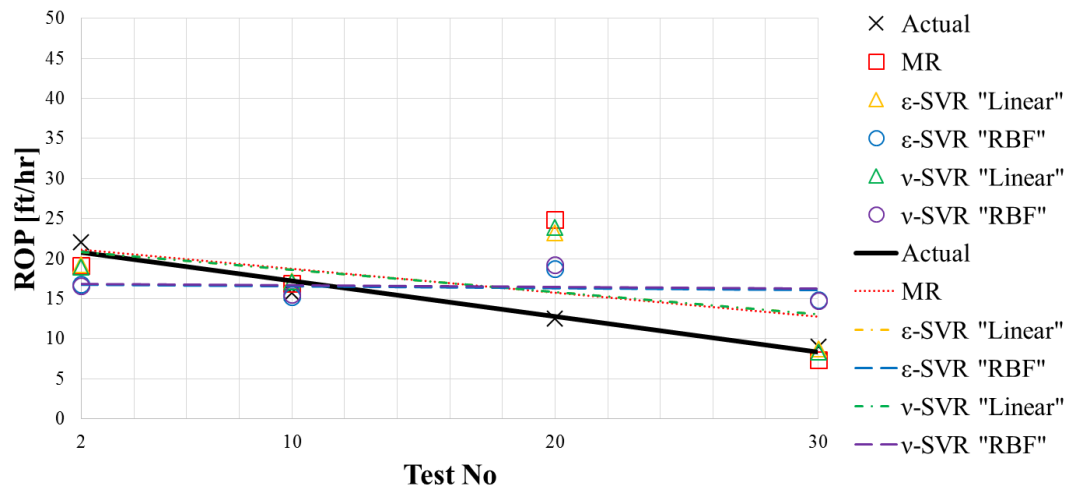


Figure 6.43 : ROP prediction trends in Case 3 for testing #2-10-20-30.

The comparison of each prediction with actual ROP is plotted in Figure 6.44.

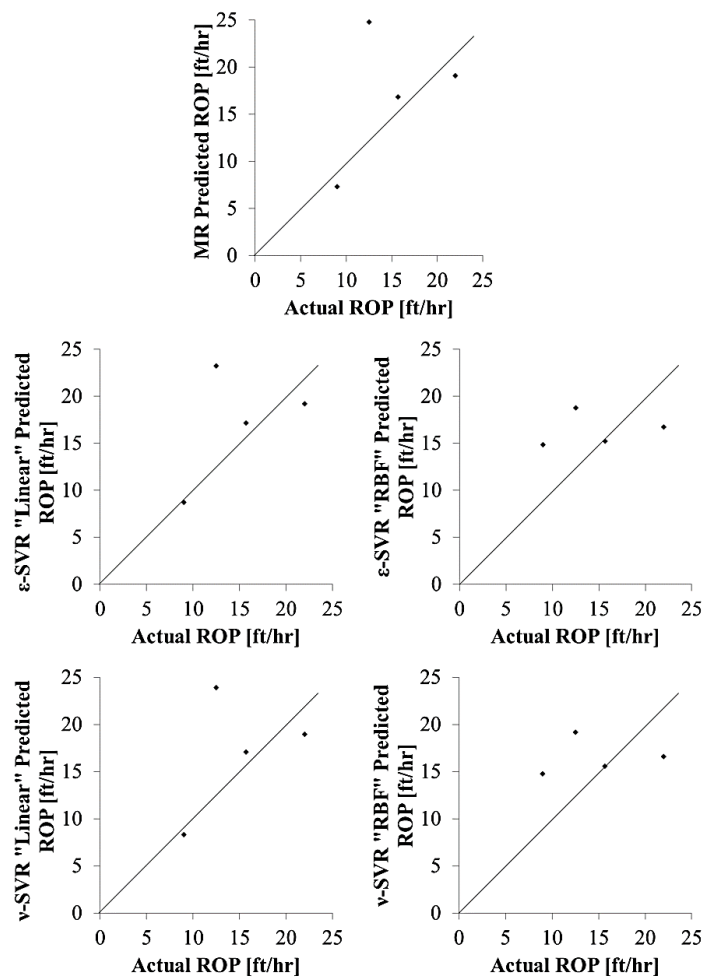


Figure 6.44 : ROP prediction comparison in Case 3 for testing #2-10-20-30.

For testing #2-10-20-30, ε -SVR with linear kernel can be considered as the best predictor. Likewise, ε -SVR with linear kernel has all the minimum errors and maximum Pseudo- R^2 value. Additionally, the best predictors are seemed to be linear kernel methods rather than MR. Moreover, RBF kernels do not seemed to be suitable for this case as good predictive methods.

The cumulative comparative statistical results for Case 3 are placed on several charts. Negative values are excluded from the graphics. The comparison of the scenarios in Case 3 in terms of RSS_{model} is shown in Figure 6.45.

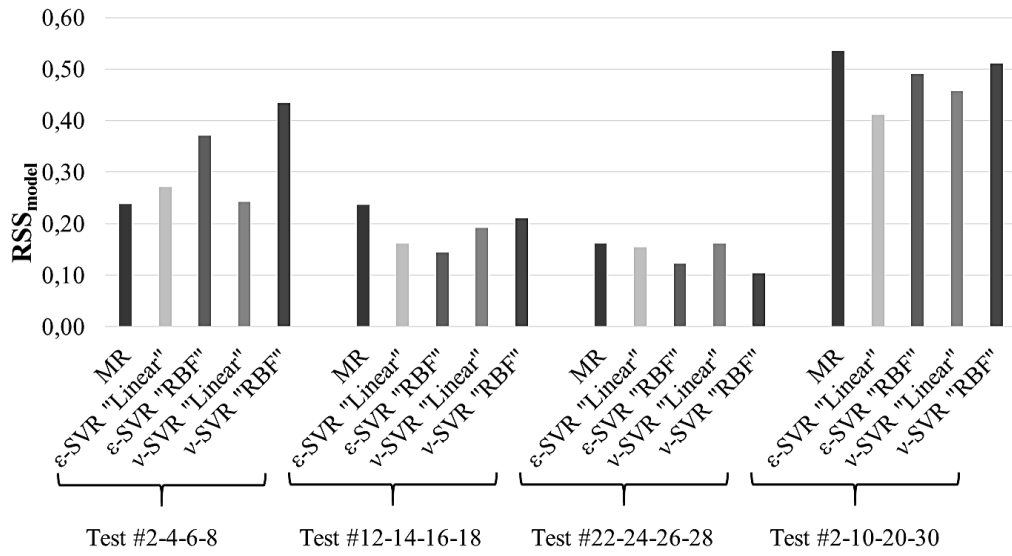


Figure 6.45 : RSS_{model} values for Case 3.

It can be seen from Figure 6.45 that MR has the maximum RSS values in most of the scenarios. On the contrary, linear kernels have less RSS values, especially ε -SVR with linear kernel has the minimum RSS for the most scenarios.

The comparison of the scenarios in Case 3 in terms of CV error is shown in Figure 6.46.

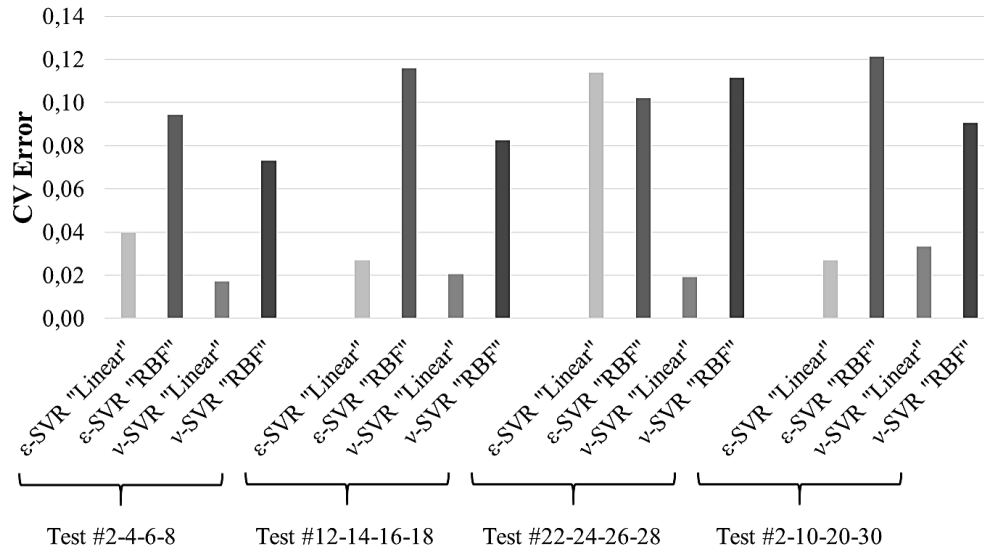


Figure 6.46 : CV error values for Case 3.

In Figure 6.46 it can be said that, in overall, SVR methods with linear kernel have the best cross validation results.

The testing errors of Case 3 in terms of RMSE is given in Figure 6.47.

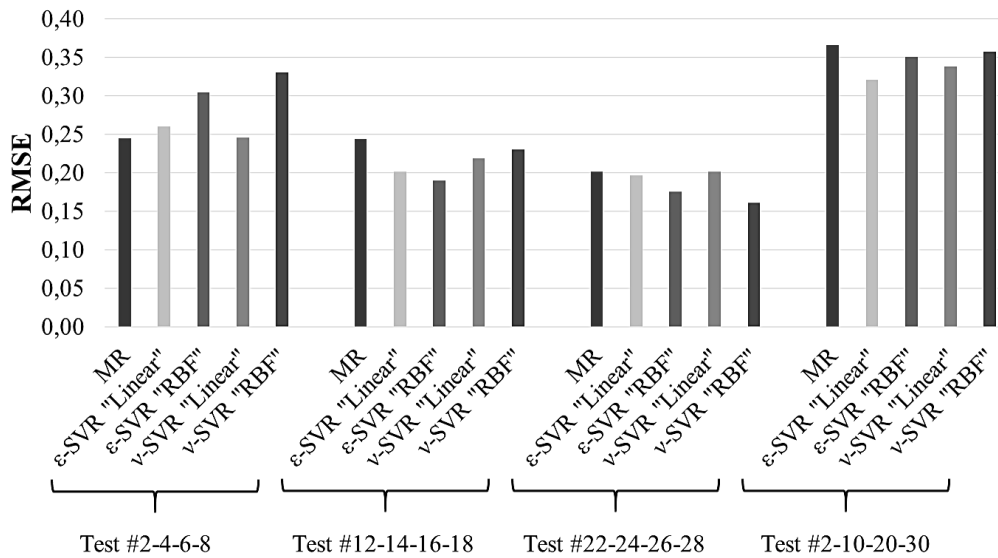


Figure 6.47 : Testing errors for Case 3.

As shown in Figure 6.47, SVR methods with linear kernel have the minimum testing error for the most of the scenarios.

The Pseudo- R^2 comparison for Case 3 is given in Figure 6.48.

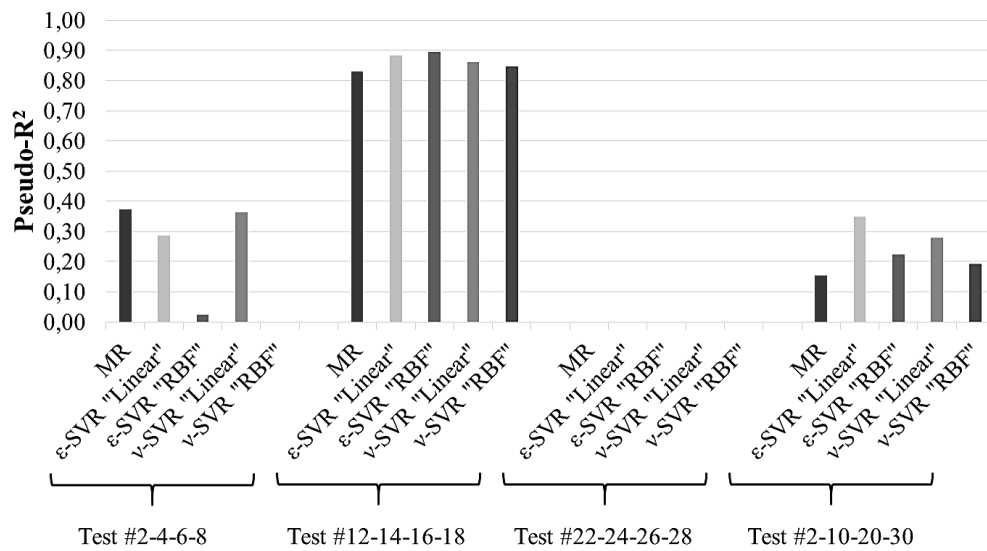


Figure 6.48 : Pseudo- R^2 values for Case 3.

It can be understood from Figure 6.48 that ϵ -SVR with linear kernel and MR can be preferred as a predicting method for most of the scenarios. However, there are negative results, which means the predicting model is not good enough. The results seems unstable. Predicting deep data inputs do not give satisfactory results.

6.4 Case 4: Training Even Numbered Data Points

In Case 4, the even numbered data points of Table A.1 are used as training data set. The coefficients of MR analysis for this training data are given in Table 6.52.

Table 6.52 : Multiple regression coefficients of Case 4.

a1	a2	a3	a4	a5	a6	a7	a8
4,133E+00	1,600E-04	1,394E-04	3,835E-05	7,668E-01	3,250E-01	4,638E-01	2,327E-01

6.4.1 Testing #1-3-5-7

In this testing scenario, the ROP values of the inputs numbered #1-3-5-7 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.49.

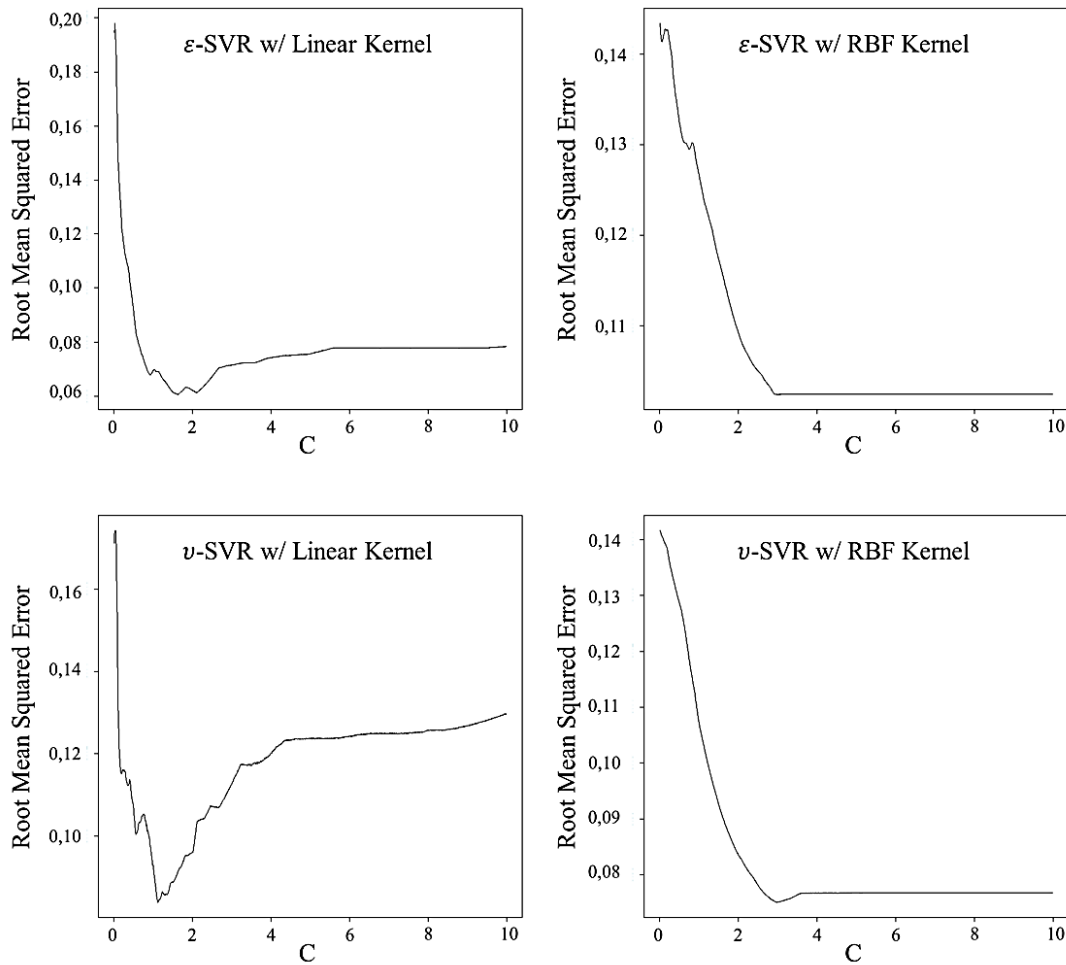


Figure 6.49 : Cost graphics of Case 4 for testing #1-3-5-7.

The basic statistical analysis of the whole data set is given in Table 6.53.

Table 6.53 : Statistics of the data set in Case 4 for testing #1-3-5-7.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,792	-2546	2578	-10588	-0,750	-0,214	-0,421	-0,231
Median	2,754	-1775	709	-7833	-0,678	-0,211	-0,380	-0,070
Variance	0,141	7863902	7682863	54500063	0,192	0,048	0,067	0,141

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.54.

Table 6.54 : Statistical results of MR in Case 4 for learning and testing #1-3-5-7.

RSS _{default}	R ²	Adjusted R ²	p-value
0,19	0,89	0,77	0,008

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.55.

Table 6.55 : Statistical results of the methods in Case 4 for testing #1-3-5-7.

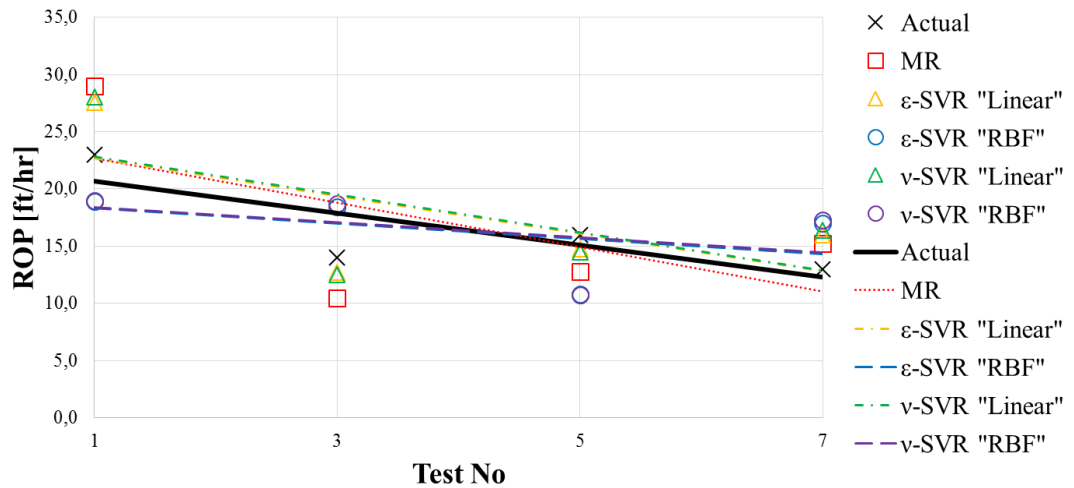
	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,21	0,09	0,34	0,11	0,36
Pseudo-R ²	-0,09	0,52	-0,76	0,41	-0,87
C		1,63	3,01	1,11	2,97
# of SV		14	14	12	15
CV Error		0,06	0,10	0,08	0,08
RMSE	0,23	0,15	0,29	0,17	0,30

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.56. The ROP values are in unit of [ft/hr].

Table 6.56 : ROP predictions in Case 4 for #1-3-5-7.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
1	29,0	27,6	18,9	28,0	19,0	23,0
3	10,5	12,7	18,5	12,5	18,7	14,0
5	12,8	14,7	10,8	14,5	10,7	16,0
7	15,2	16,0	17,0	16,4	17,3	13,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.50.

**Figure 6.50** : ROP prediction trends in Case 4 for testing #1-3-5-7.

The comparison of each prediction with actual ROP is plotted in Figure 6.51.

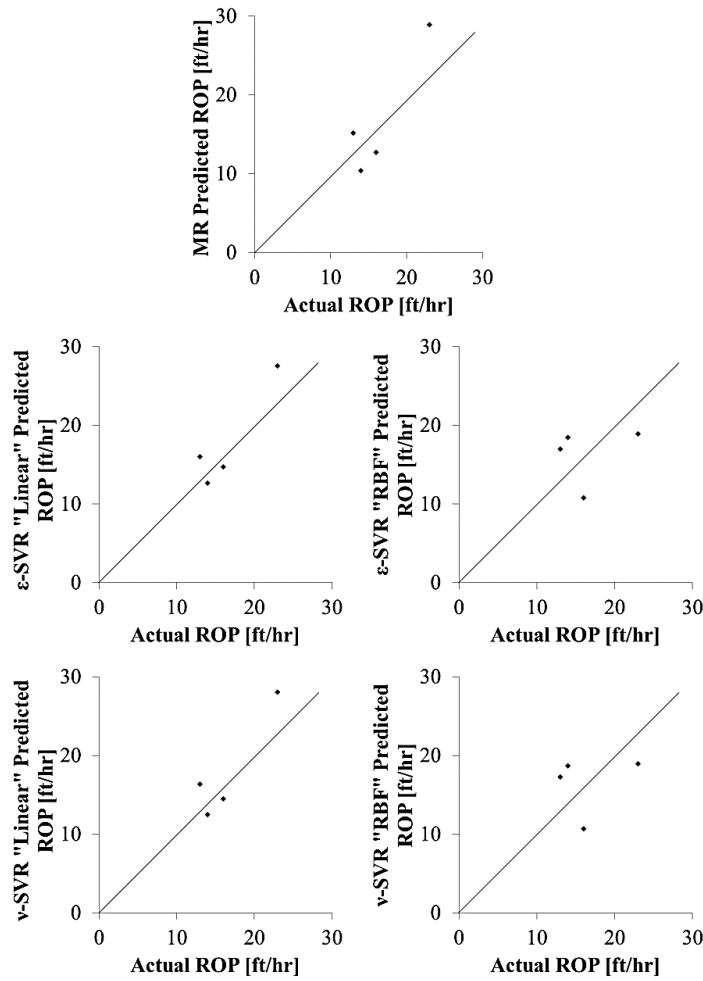


Figure 6.51 : ROP prediction comparison in Case 4 for testing #1-3-5-7.

It can be seen from Figure 6.51 and 6.50 that the best predictive methods in Case 4 for testing #1-3-5-7 are the SVR methods with linear kernel, especially ε -SVR, since they have the minimum RSS and RMSE and maximum Pseudo- R^2 ratio among the other methods.

6.4.2 Testing #11-13-15-17

In this testing scenario, the ROP values of the inputs numbered #11-13-15-17 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.52.

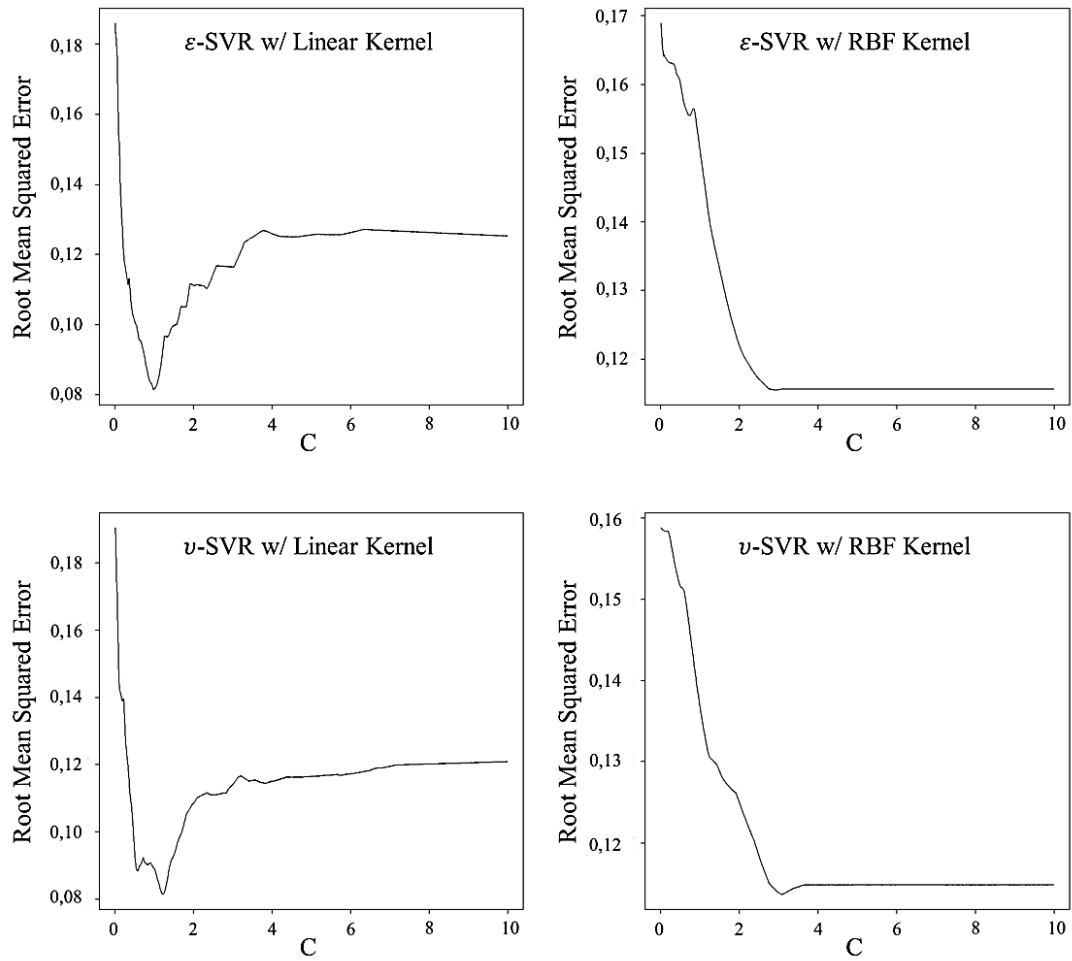


Figure 6.52 : Cost graphics of Case 4 for testing #11-13-15-17.

The basic statistical analysis of the whole data set is given in Table 6.57.

Table 6.57 : Statistics of the data set in Case 4 for testing #11-13-15-17.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,784	-2972	3151	-13304	-0,617	-0,277	-0,342	-0,206
Median	2,741	-2315	3588	-10118	-0,571	-0,261	-0,330	-0,070
Variance	0,236	6460831	6537582	83289886	0,142	0,029	0,049	0,156

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.58.

Table 6.58 : Statistical results of MR in Case 4 for testing #11-13-15-17.

RSS _{default}	R ²	Adjusted R ²	p-value
1,89	0,89	0,77	0,008

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.59.

Table 6.59 : Statistical results of the methods in Case 4 for testing #11-13-15-17.

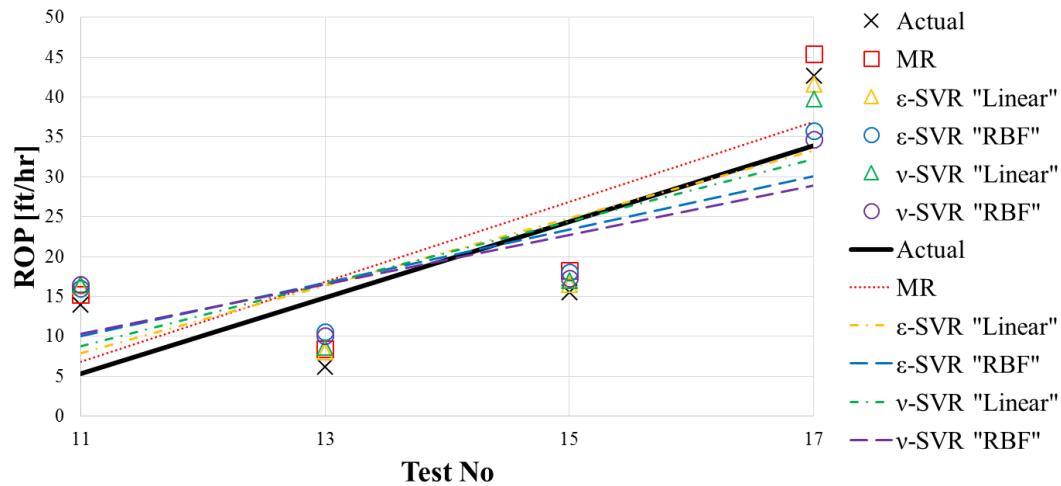
	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,13	0,09	0,36	0,15	0,32
Pseudo-R ²	0,93	0,95	0,81	0,92	0,83
C		0,98	2,925	1,22	3,08
# of SV		14	14	13	15
CV Error		0,08	0,12	0,08	0,11
RMSE	0,18	0,15	0,30	0,19	0,28

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.60. The ROP values are in unit of [ft/hr].

Table 6.60 : ROP predictions in Case 4 for #11-13-15-17.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
11	15,2	16,3	16,0	16,5	16,5	14,0
13	8,5	8,0	10,6	8,6	10,1	6,2
15	18,3	16,5	18,1	17,0	17,3	15,5
17	45,4	41,6	35,7	39,8	34,7	42,7

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.53.

**Figure 6.53** : ROP prediction trends in Case 4 for testing #11-13-15-17.

The comparison of each prediction with actual ROP is plotted in Figure 6.54.

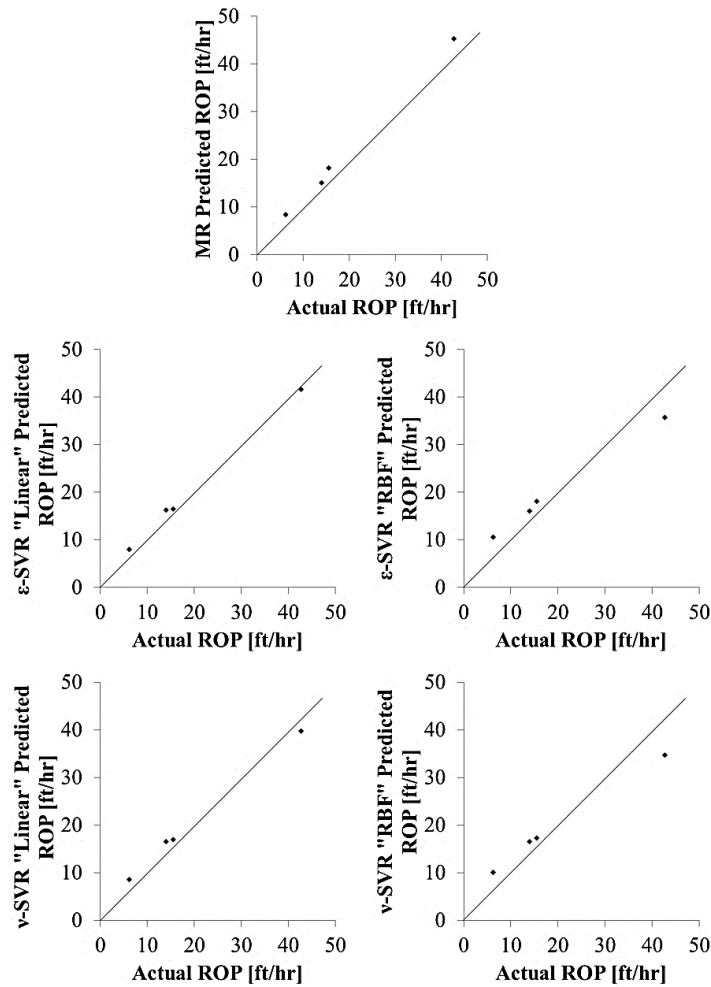


Figure 6.54 : ROP prediction comparison in Case 4 for testing #11-13-15-17.

For this scenario, the maximum correlation seems to be ε -SVR with linear kernel from Figure 6.54 and 6.53. Besides, ε -SVR with linear kernel has the maximum Pseudo- R^2 value, the minimum errors, and the minimum RSS_{model} value. Thus, ε -SVR with linear kernel is the best predicting method.

6.4.3 Testing #21-23-25-27

In this testing scenario, the ROP values of the inputs numbered #21-23-25-27 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.55.

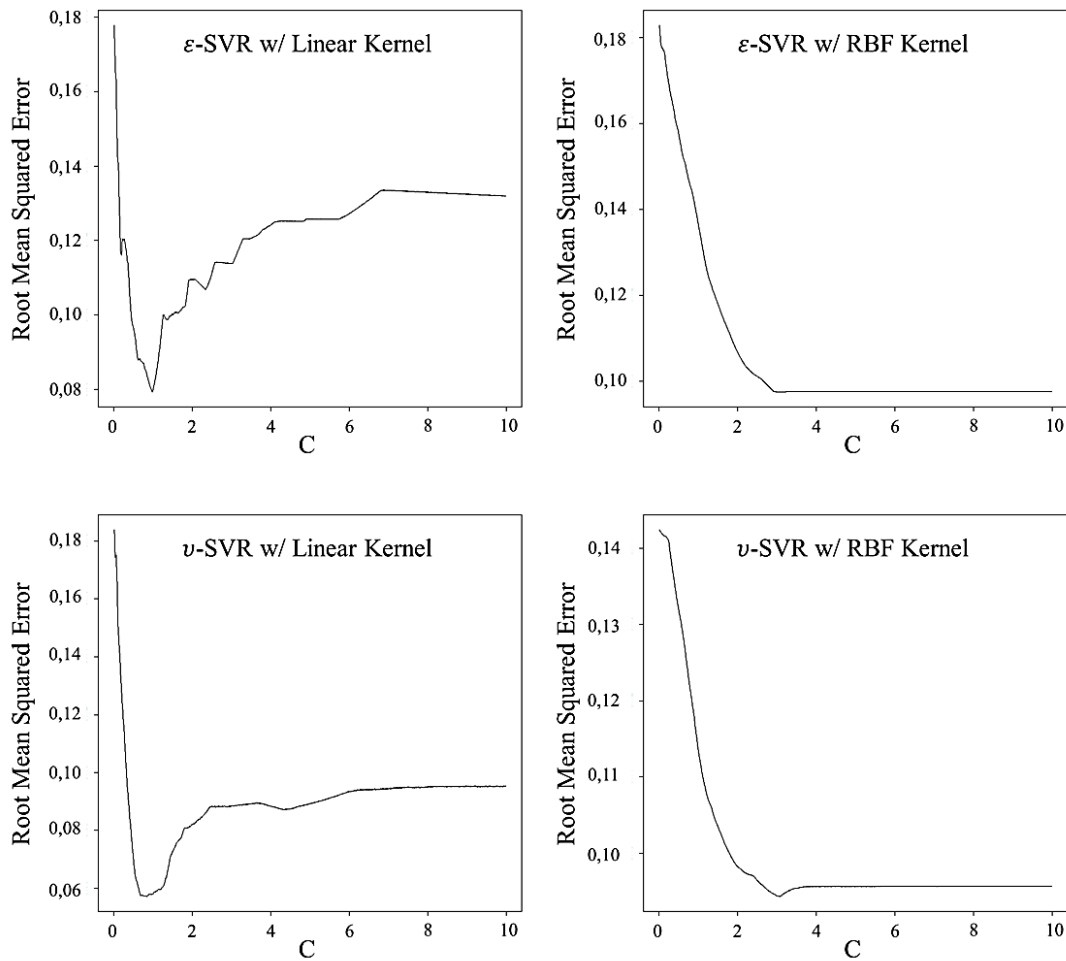


Figure 6.55 : Cost graphics of Case 4 for testing #21-23-25-27.

The basic statistical analysis of the whole data set is given in Table 6.61.

Table 6.61 : Statistics of the data set in Case 4 for testing #21-23-25-27.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,829	-3578	3753	-12090	-0,687	-0,259	-0,352	-0,320
Median	2,809	-3795	5173	-10118	-0,659	-0,249	-0,330	-0,160
Variance	0,152	6967823	6846027	56090453	0,144	0,028	0,049	0,245

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.62.

Table 6.62 : Statistical results of MR in Case 4 for learning and testing #21-23-25-27.

RSS _{default}	R ²	Adjusted R ²	p-value
0,40	0,89	0,77	0,008

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.63.

Table 6.63 : Statistical results of the methods in Case 4 for testing #21-23-25-27.

	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,78	0,69	0,38	0,93	0,46
Pseudo-R ²	-0,92	-0,71	0,06	-1,29	-0,13
C		0,98	3,07	0,845	3,07
# of SV		14	14	12	15
CV Error		0,08	0,10	0,06	0,09
RMSE	0,44	0,42	0,31	0,48	0,34

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.64. The ROP values are in unit of [ft/hr].

Table 6.64 : ROP predictions in Case 4 for #21-23-25-27.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
21	8,9	9,2	12,0	8,1	11,2	21,1
23	17,0	17,8	22,9	17,2	22,6	18,7
25	23,7	26,6	26,2	25,6	25,3	27,1
27	12,5	12,8	14,4	12,7	14,1	12,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.56.

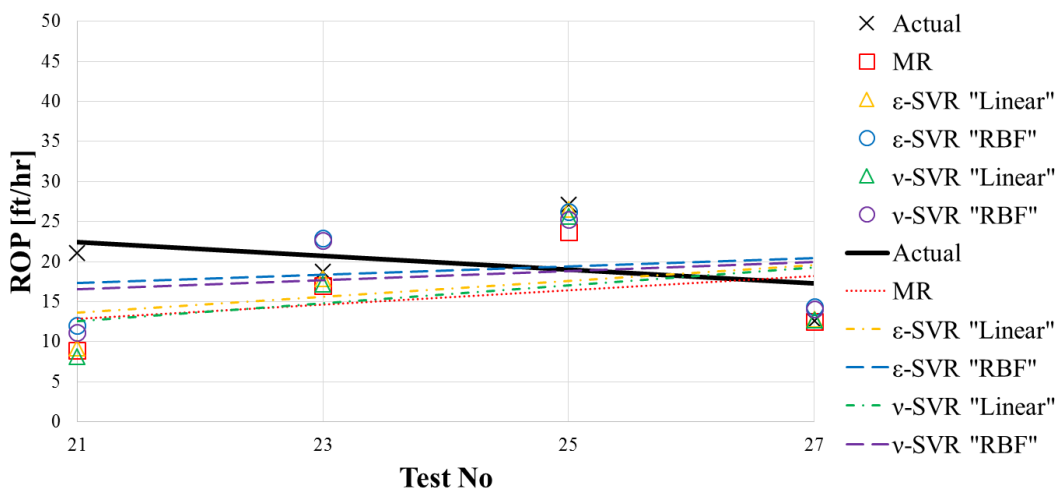


Figure 6.56 : ROP prediction trends in Case 4 for testing #21-23-25-27.

The comparison of each prediction with actual ROP is plotted in Figure 6.57.

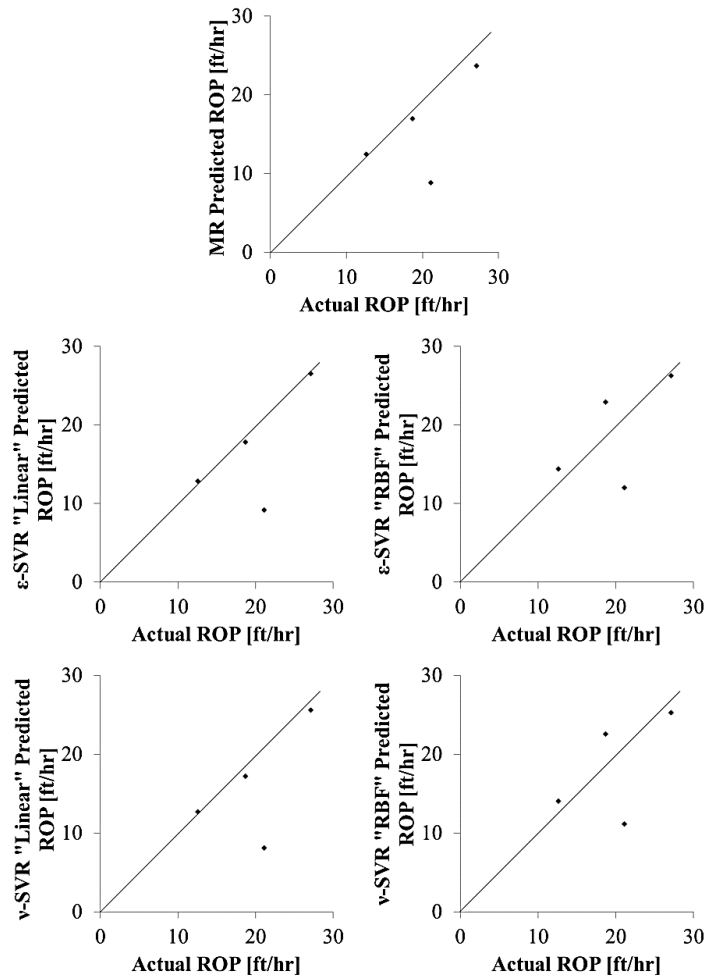


Figure 6.57 : ROP prediction comparison in Case 4 for testing #21-23-25-27.

It can be understood from Figure 6.57 that, SVR methods with linear kernel seemed to be the best predictors. However, in Figure 6.56 it can be seen that none of the methods have good approximation to the prediction model. Because, the residual of one data point is generating high deviation. The reason for that situation is the insufficient input number for the learning process.

6.4.4 Testing #1-9-19-29

In this testing scenario, the ROP values of the inputs numbered #1-9-19-29 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.58.

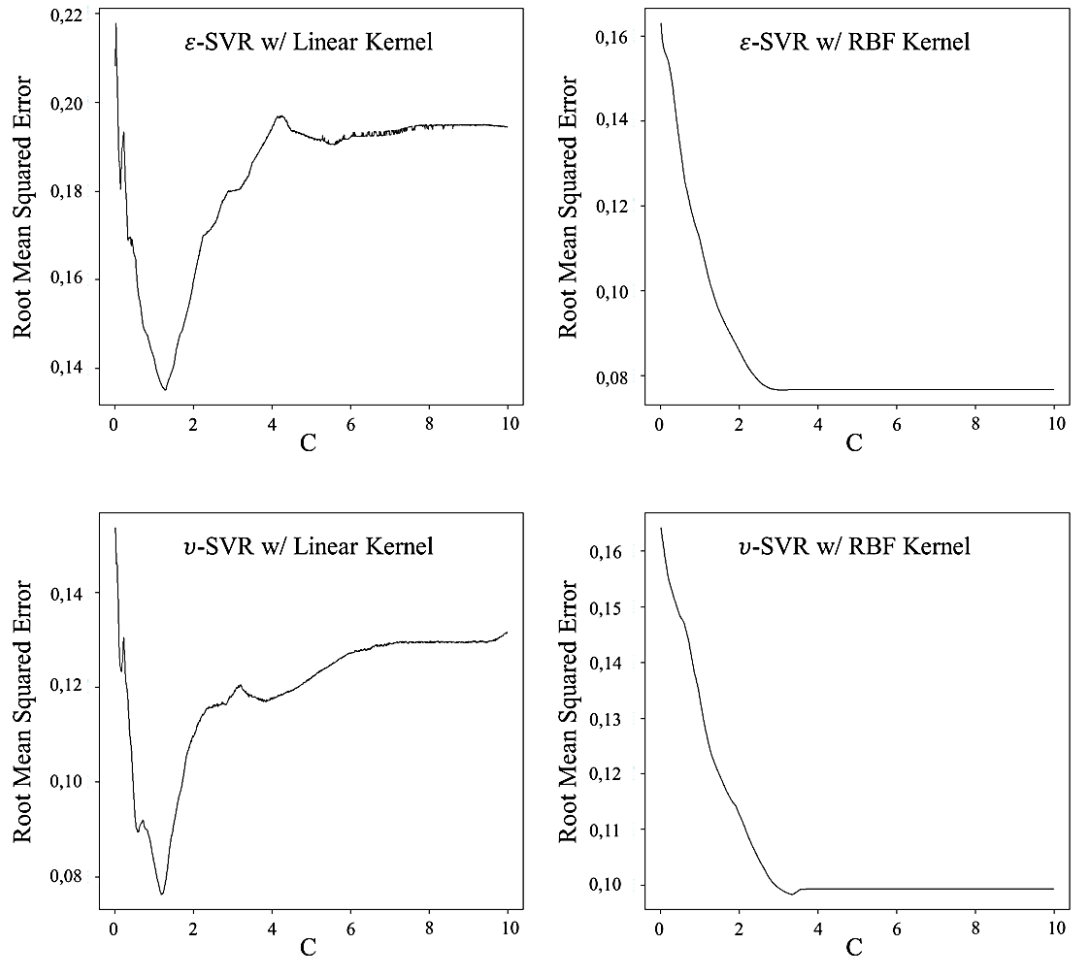


Figure 6.58 : Cost graphics of Case 4 for testing #1-9-19-29.

The basic statistical analysis of the whole data set is given in Table 6.65.

Table 6.65 : Statistics of the data set in Case 4 for testing #1-9-19-29.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,854	-3082	3183	-11154	-0,607	-0,263	-0,432	-0,234
Median	2,766	-2900	4733	-8277	-0,549	-0,274	-0,420	-0,130
Variance	0,188	8062064	7638290	55711976	0,143	0,039	0,068	0,145

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.66.

Table 6.66 : Statistical results of MR in Case 4 for learning and testing #1-9-19-29.

RSS _{default}	R ²	Adjusted R ²	p-value
1,10	0,89	0,77	0,008

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.67.

Table 6.67 : Statistical results of the methods in Case 4 for testing #1-9-19-29.

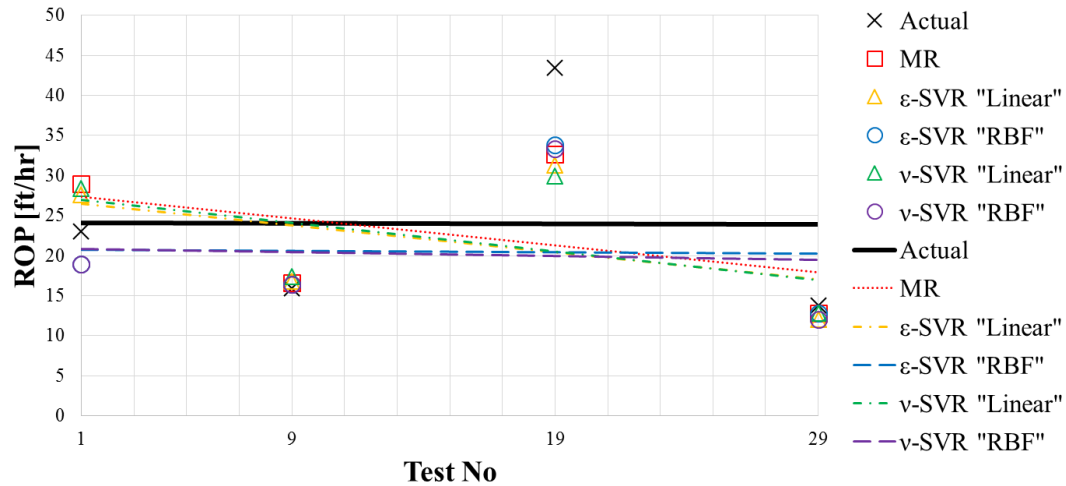
	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,14	0,16	0,11	0,20	0,13
Pseudo-R ²	0,87	0,85	0,90	0,82	0,88
C		1,27	3,095	1,185	3,35
# of SV		14	14	13	15
CV Error		0,14	0,08	0,08	0,10
RMSE	0,19	0,20	0,16	0,22	0,18

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.68. The ROP values are in unit of [ft/hr].

Table 6.68 : ROP predictions in Case 4 for #1-9-19-29.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
1	29,0	27,6	18,9	28,3	18,9	23,0
9	16,6	16,8	16,4	17,4	16,4	15,9
19	32,7	31,3	33,8	29,9	33,3	43,4
29	12,8	12,0	12,8	12,8	12,0	13,8

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.59.

**Figure 6.59** : ROP prediction trends in Case 4 for testing #1-9-19-29.

The comparison of each prediction with actual ROP is plotted in Figure 6.60.

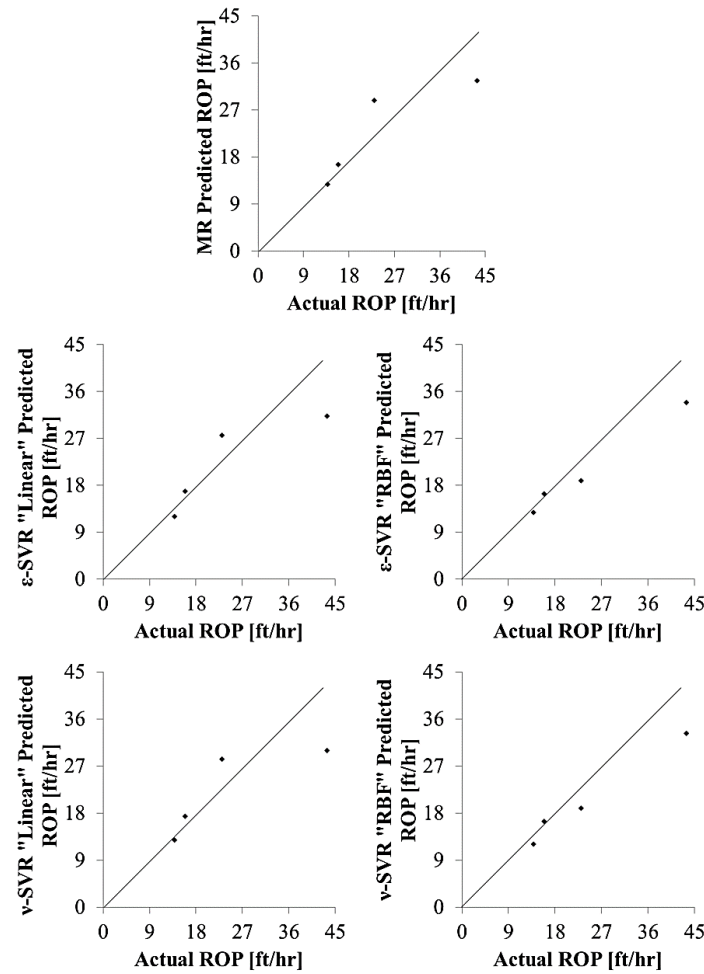


Figure 6.60 : ROP prediction comparison in Case 4 for testing #1-9-19-29.

For testing #1-9-19-29, SVR methods with RBF kernel can be considered as the best predictors. Likewise, ε -SVR with RBF kernel has all the minimum errors and maximum Pseudo- R^2 value. Additionally, ν -SVR with RBF kernel has the minimum RSS_{model} value.

The cumulative comparative statistical results for Case 4 are placed on several charts. Negative values are excluded from the graphics. The comparison of the scenarios in Case 4 in terms of RSS_{model} is shown in Figure 6.61.

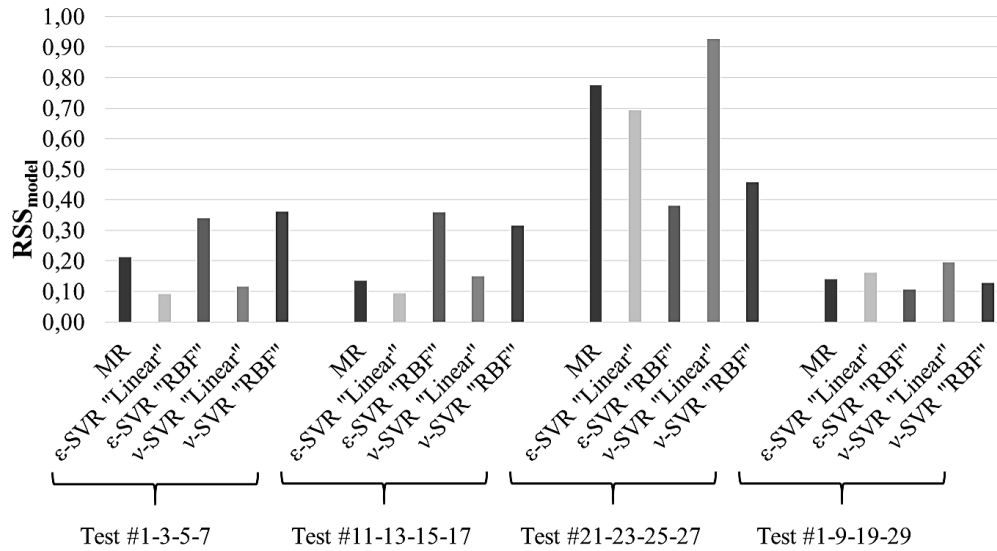


Figure 6.61 : RSS_{model} values for Case 4.

It can be seen from Figure 6.61 that SVR methods with RBF kernel have the maximum RSS values in the first two scenarios. However, the results are the exact opposite for the third and fourth scenarios.

The comparison of the scenarios in Case 4 in terms of CV error is shown in Figure 6.62.

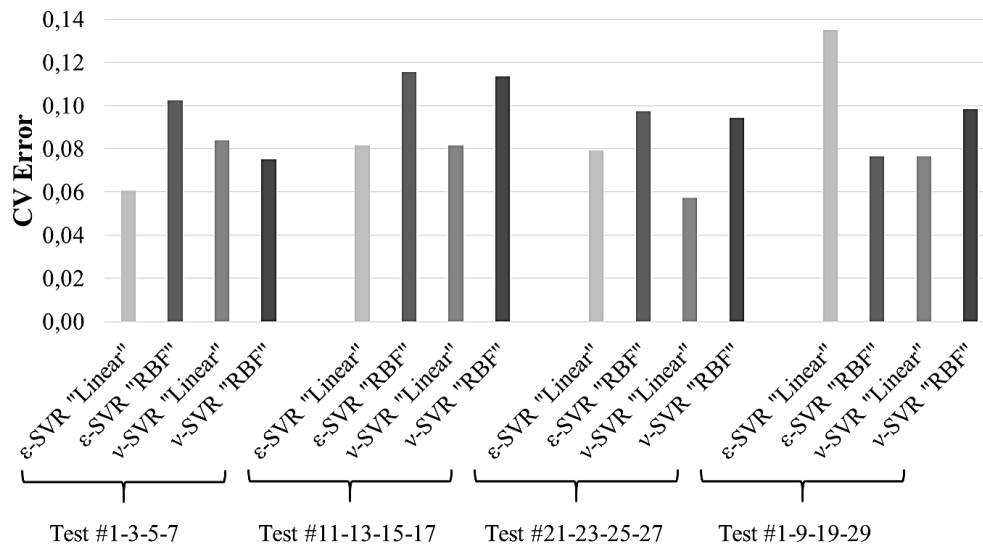


Figure 6.62 : CV error values for Case 4.

In Figure 6.62 it can be said that, the results are variable. In summary, linear kernels can be considered as better than RBF kernels in terms of CV error.

The testing errors of Case 4 in terms of RMSE is given in Figure 6.63.

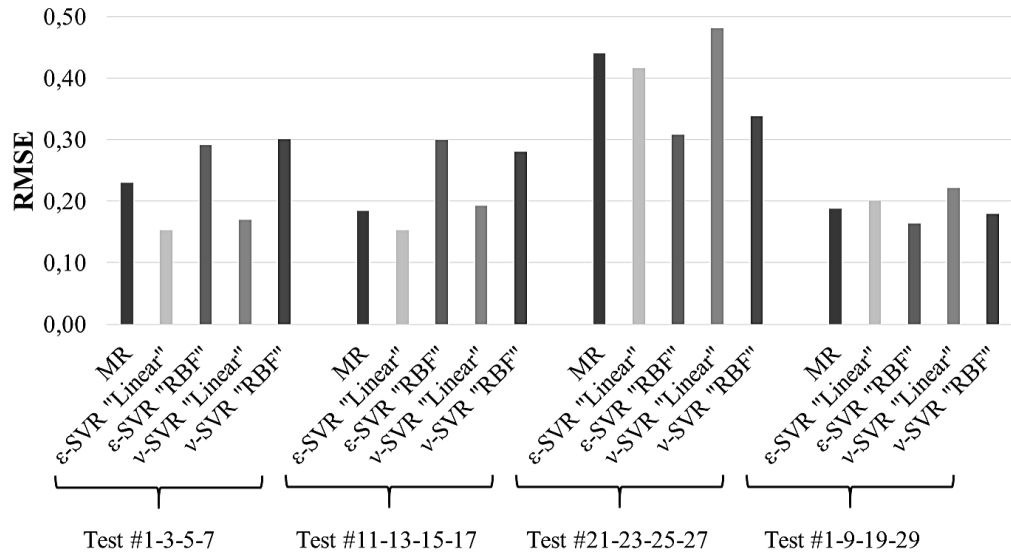


Figure 6.63 : Testing errors for Case 4.

As shown in Figure 6.63, the results are similar with Figure 6.61. SVR methods with RBF kernel have the maximum error in the first two scenarios. However, the results are total opposite for other scenarios.

The Pseudo- R^2 comparison for Case 4 is given in Figure 6.64.

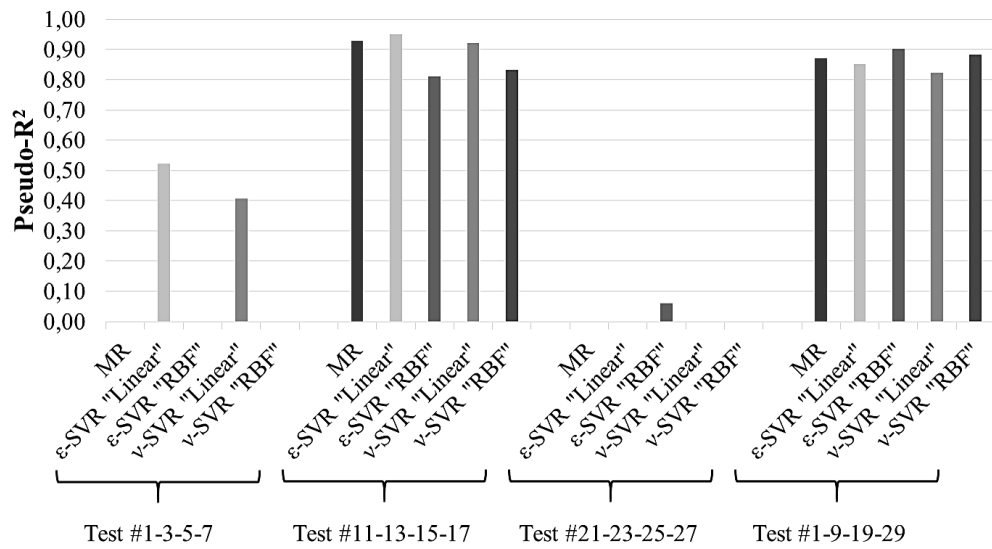


Figure 6.64 : Pseudo- R^2 values for Case 4.

It can be understood from Figure 6.64 that ϵ -SVR with linear kernel can be preferred as a predicting method for most of the scenarios. However, same as the previous case, there are negative results, which means the models are not good enough to make accurate predictions.

6.5 Case 5: Training Several 24 Data Points

In Case 5, several 24 data points of Table A.1 are used as training data set.

6.5.1 Training first 24, testing last 6 data inputs

In this testing scenario, the ROP values of the last 6 data inputs are desired to be predicted. The coefficients of MR analysis for this training data are given in Table 6.69.

Table 6.69 : Multiple regression coefficients in Case 5 for testing last 6.

a1	a2	a3	a4	a5	a6	a7	a8
3,851E+00	1,536E-04	1,827E-04	4,302E-05	4,815E-01	3,086E-01	4,408E-01	1,273E-01

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.65.

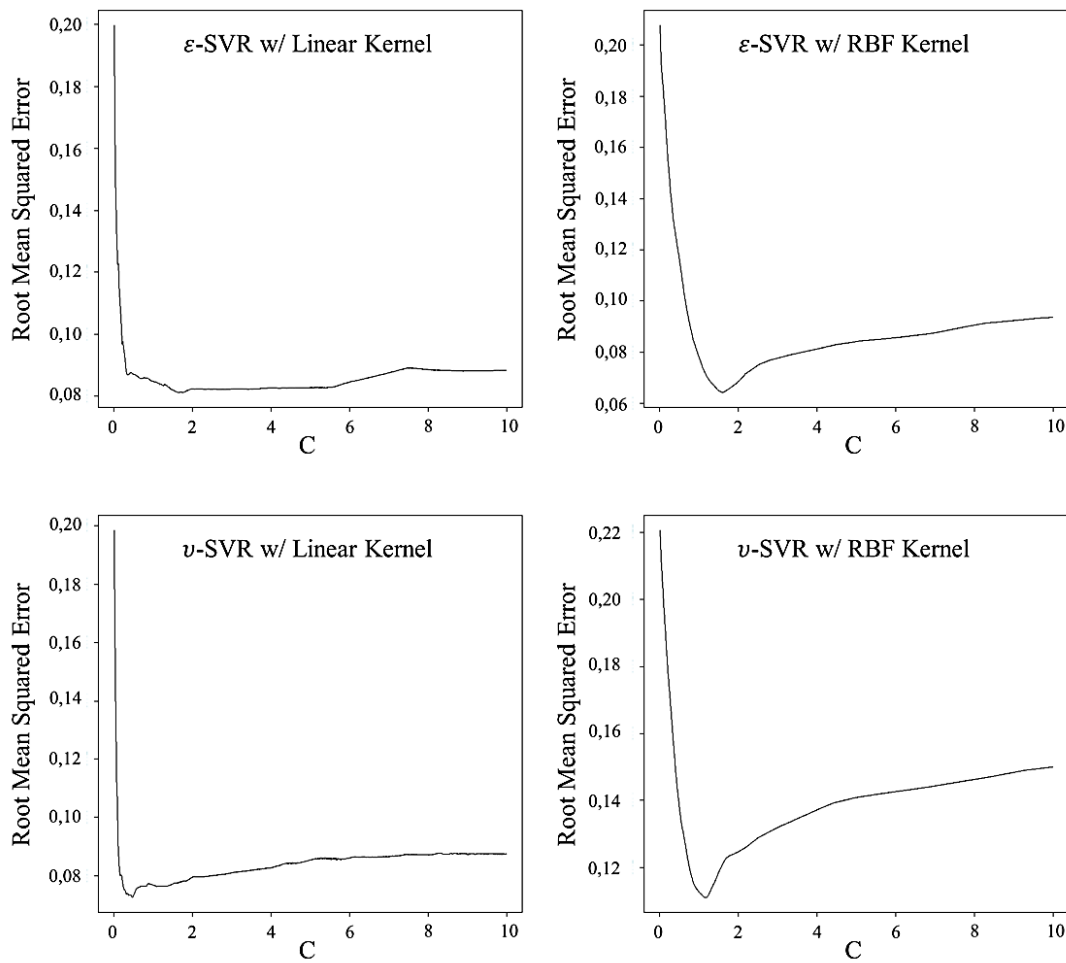


Figure 6.65 : Cost graphics of Case 5 for testing last 6.

The basic statistical analysis of the whole data set is given in Table 6.70.

Table 6.70 : Statistics of the data set in Case 5 for testing last 6.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,833	-2952	3135	-11916	-0,680	-0,248	-0,401	-0,213
Median	2,760	-2608	4160	-8342	-0,647	-0,261	-0,380	-0,095
Variance	0,199	6530787	7008892	66623679	0,174	0,035	0,060	0,181

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.71.

Table 6.71 : Statistical results of MR in Case 5 for learning and testing last 6.

RSS _{default}	R ²	Adjusted R ²	p-value
0,87	0,85	0,78	1,90E-05

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.72.

Table 6.72 : Statistical results of the methods in Case 5 for testing last 6.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,10	0,08	0,78	1,15	0,72
Pseudo-R ²	0,88	0,91	0,11	-0,31	0,17
C		1,75	1,605	1,11	1,17
# of SV		19	18	15	18
CV Error		0,08	0,06	0,08	0,11
RMSE	0,13	0,12	0,36	0,44	0,35

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.73. The ROP values are in unit of [ft/hr].

Table 6.73 : ROP predictions in Case 5 for last 6.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
25	27,1	27,5	24,6	32,2	24,7	27,1
26	16,3	15,2	18,8	20,1	18,8	14,8
27	13,8	13,8	16,6	17,3	17,1	12,6
28	19,5	18,2	20,6	24,7	19,4	14,9
29	13,3	11,9	18,3	18,6	18,0	13,8
30	9,9	9,9	17,5	19,2	17,3	9,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.66.

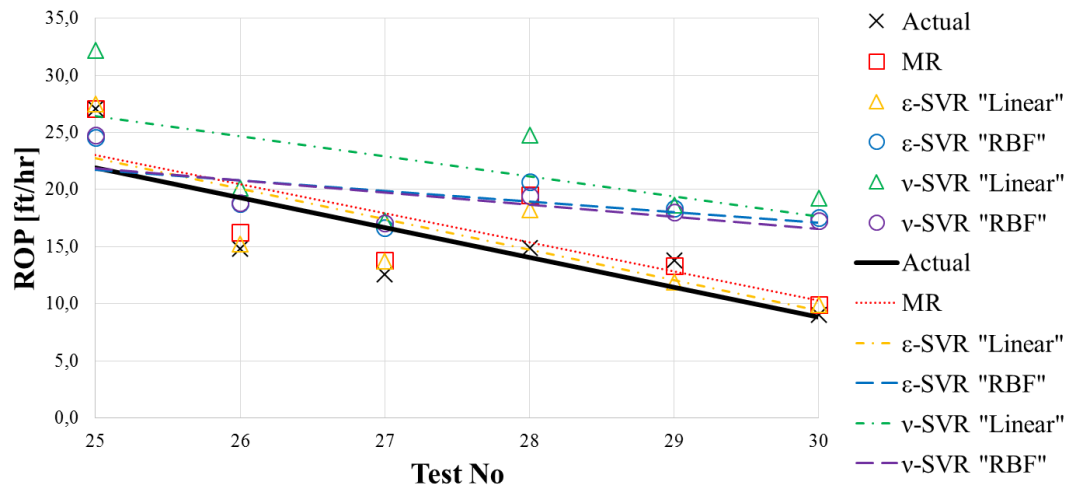


Figure 6.66 : ROP prediction trends in Case 5 for testing last 6.

The comparison of each prediction with actual ROP is plotted in Figure 6.67.

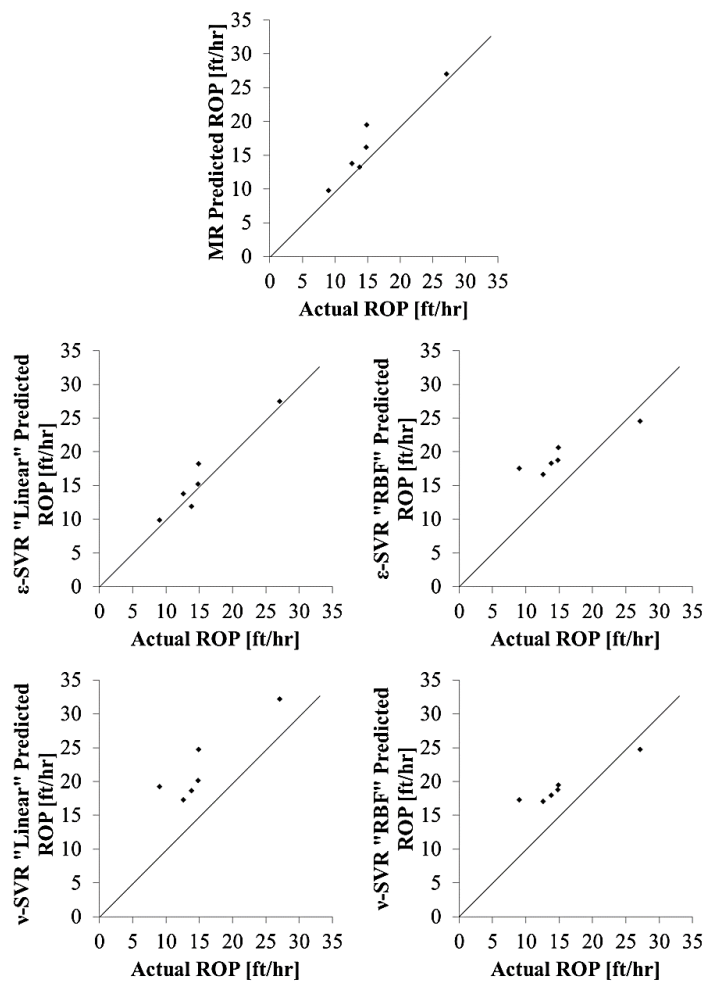


Figure 6.67 : ROP prediction comparison in Case 5 for testing last 6.

It can be seen from Figure 6.67 and 6.66 that the best predictive methods in Case 5 for testing last 6 are MR and ε -SVR with linear kernel, since they have the minimum RSS and RMSE and maximum Pseudo- R^2 ratio among the other methods.

6.5.2 Training last 24, testing first 6 data inputs

In this testing scenario, the ROP values of the first 6 data inputs are desired to be predicted. The coefficients of MR analysis for this training data are given in Table 6.74.

Table 6.74 : Multiple regression coefficients in Case 5 for testing first 6.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
3,586E+00	1,909E-04	2,265E-04	4,371E-05	3,484E-01	-2,815E-01	5,072E-01	1,948E-01

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.68.

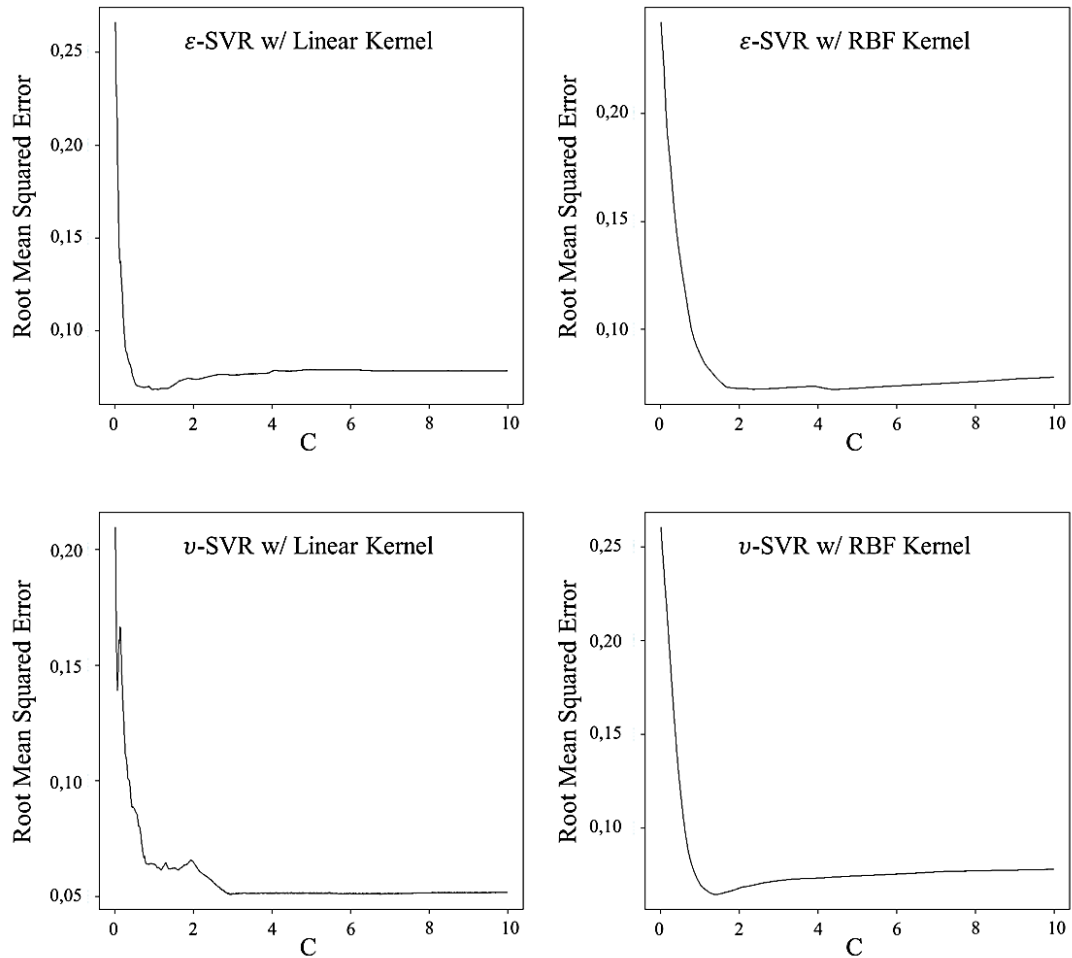


Figure 6.68 : Cost graphics of Case 5 for testing first 6.

The basic statistical analysis of the whole data set is given in Table 6.75.

Table 6.75 : Statistics of the data set in Case 5 for testing first 6.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,833	-2952	3135	-11916	-0,680	-0,248	-0,401	-0,213
Median	2,760	-2608	4160	-8342	-0,647	-0,261	-0,380	-0,095
Variance	0,199	6530787	7008892	66623679	0,174	0,035	0,060	0,181

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.76.

Table 6.76 : Statistical results of MR in Case 5 for testing first 6.

RSS _{default}	R ²	Adjusted R ²	p-value
0,49	0,90	0,86	4,76E-07

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.77.

Table 6.77 : Statistical results of the methods in Case 5 for testing first 6.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,59	0,29	0,64	0,64	0,65
Pseudo-R ²	-0,19	0,40	-0,30	-0,29	-0,31
C		0,94	4,37	2,915	1,37
# of SV		16	21	14	21
CV Error		0,07	0,07	0,05	0,06
RMSE	0,31	0,22	0,33	0,33	0,33

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.78. The ROP values are in unit of [ft/hr].

Table 6.78 : ROP predictions in Case 5 for first 6.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
1	17,9	23,1	14,2	17,2	14,2	23,0
2	14,9	17,4	13,9	14,5	13,9	22,0
3	9,5	11,1	13,9	9,4	14,0	14,0
4	14,6	13,9	12,9	14,7	12,7	10,0
5	14,3	13,3	12,8	14,5	13,1	16,0
6	14,5	15,4	14,3	14,6	13,7	19,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.69.

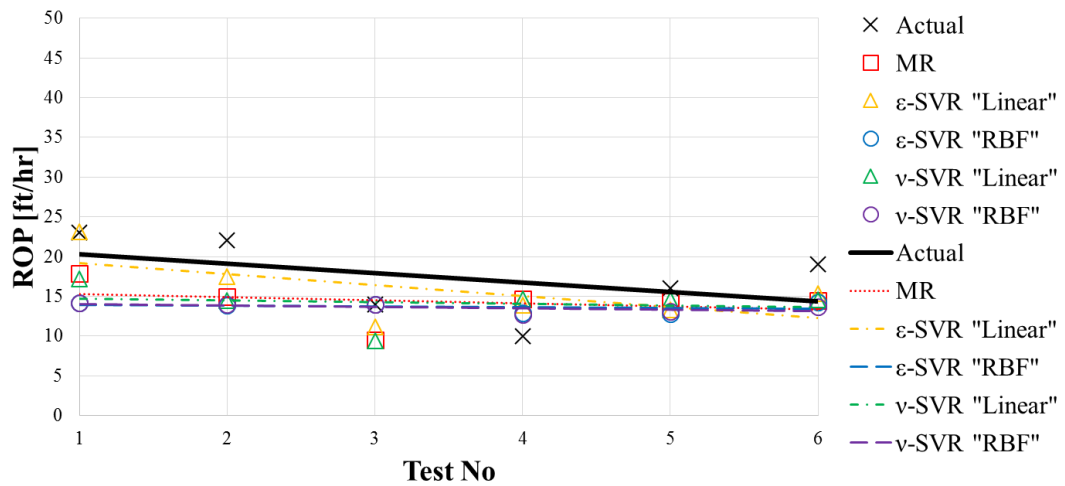


Figure 6.69 : ROP prediction trends in Case 5 for testing first 6.

The comparison of each prediction with actual ROP is plotted in Figure 6.70.

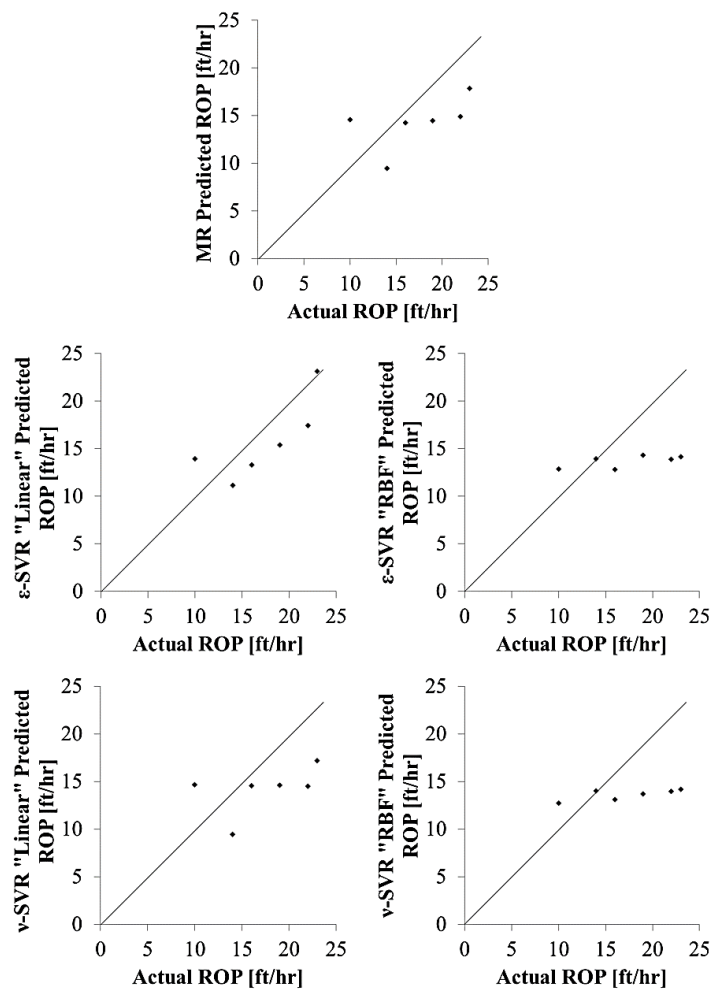


Figure 6.70 : ROP prediction comparison in Case 5 for testing first 6.

For this scenario, the best approach is seems to be in ε -SVR with linear kernel, which can be understood from Figure 6.70 and 6.69. Besides, ε -SVR with linear kernel has the maximum Pseudo- R^2 value, the minimum errors, and the minimum RSS_{model} value.

6.5.3 Training mid 24, testing #1-2-3-28-29-30

In this testing scenario, the ROP values of the inputs numbered #1-2-3-28-29-30 are desired to be predicted. The coefficients of MR analysis for this training data are given in Table 6.79.

Table 6.79 : Multiple regression coefficients in Case 5 for testing #1-2-3-28-29-30.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
3,774E+00	1,775E-04	2,008E-04	4,465E-05	5,175E-01	-1,508E-01	4,883E-01	1,156E-01

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.71.

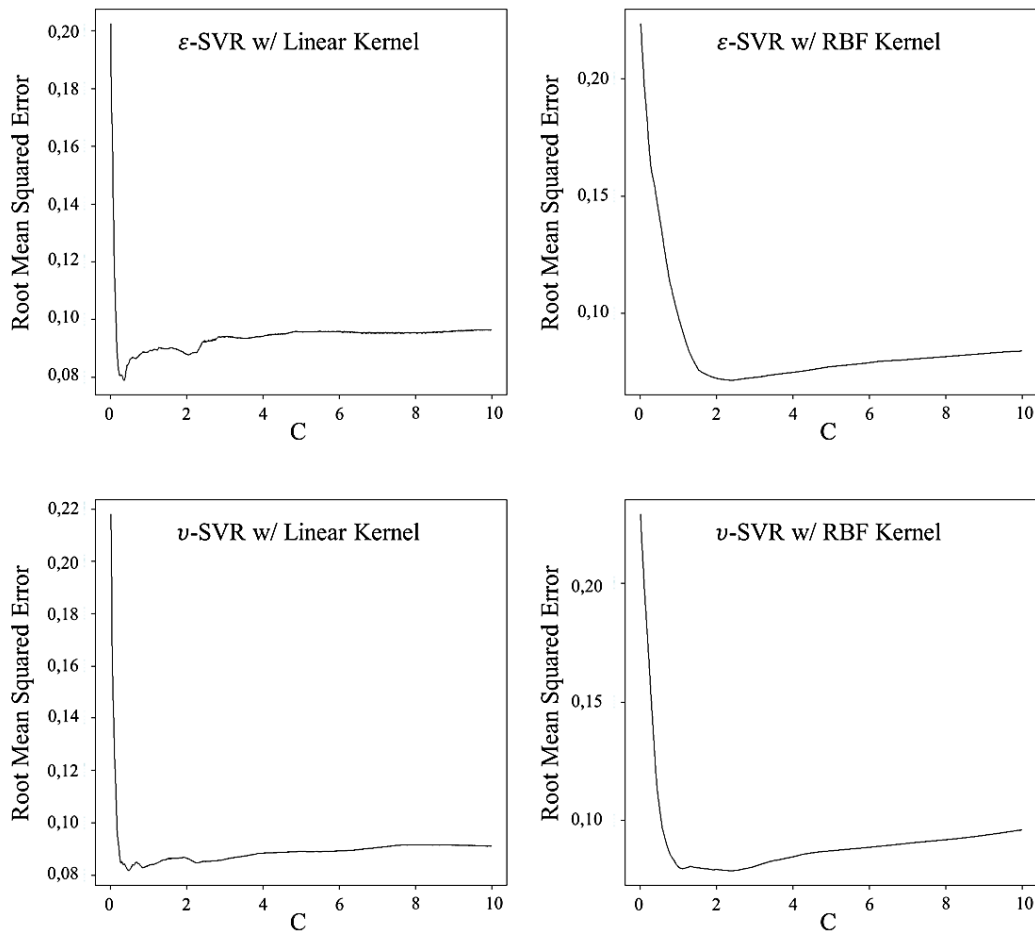


Figure 6.71 : Cost graphics of Case 5 for testing #1-2-3-28-29-30.

The basic statistical analysis of the whole data set is given in Table 6.80.

Table 6.80 : Statistics of the data set in Case 5 for testing #1-2-3-28-29-30.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,833	-2952	3135	-11916	-0,680	-0,248	-0,401	-0,213
Median	2,760	-2608	4160	-8342	-0,647	-0,261	-0,380	-0,095
Variance	0,199	6530787	7008892	66623679	0,174	0,035	0,060	0,181

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.81.

Table 6.81 : Statistical results of MR in Case 5 for learning and testing #1-2-3-28-29-30.

RSS _{default}	R ²	Adjusted R ²	p-value
0,69	0,87	0,81	6,45E-06

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.82.

Table 6.82 : Statistical results of the methods in Case 5 for testing #1-2-3-28-29-30.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,38	0,92	0,75	0,53	0,74
Pseudo-R ²	0,46	-0,33	-0,08	0,24	-0,07
C		0,36	2,405	0,48	2,35
# of SV		19	20	15	21
CV Error		0,08	0,07	0,08	0,08
RMSE	0,25	0,39	0,35	0,30	0,35

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.83. The ROP values are in unit of [ft/hr].

Table 6.83 : ROP predictions in Case 5 for #1-2-3-28-29-30.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
1	20,5	21,6	14,5	20,2	14,5	23,0
2	15,1	14,9	13,3	16,2	13,3	22,0
3	9,3	8,6	14,1	10,4	14,1	14,0
28	18,3	22,1	18,5	19,5	18,5	14,9
29	13,2	15,8	15,0	15,6	15,0	13,8
30	10,0	16,3	14,5	14,7	14,5	9,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.72.

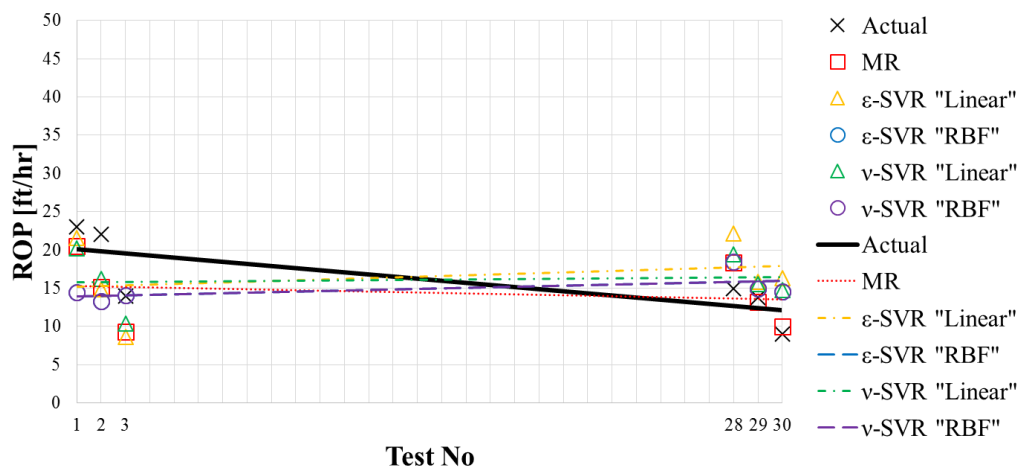


Figure 6.72 : ROP prediction trends in Case 5 for testing #1-2-3-28-29-30.

The comparison of each prediction with actual ROP is plotted in Figure 6.73.

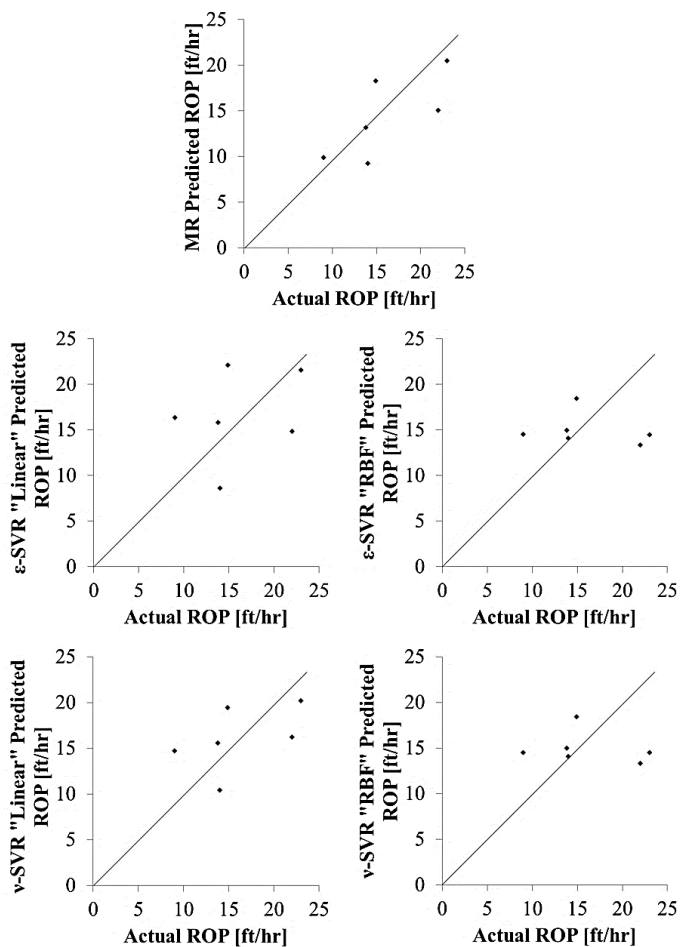


Figure 6.73 : ROP prediction comparison in Case 5 for testing #1-2-3-28-29-30.

It can be understood from Figure 6.73 and 6.72 that, MR seemed to be the best predictor method for this scenario. Moreover, MR has the maximum Pseudo- R^2 value, minimum RMSE and RSS among other methods.

6.5.4 Training several 24, testing #4-7-10-18-22-27

In this testing scenario, the ROP values of the inputs numbered #4-7-10-18-22-27 are desired to be predicted while remaining data are trained. The coefficients of MR analysis for this training data are given in Table 6.84.

Table 6.84 : Multiple regression coefficients in Case 5 for testing #4-7-10-18-22-27.

a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
3,717E+00	1,701E-04	2,006E-04	4,528E-05	2,710E-01	4,090E-02	4,611E+00	2,365E-01

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.74.

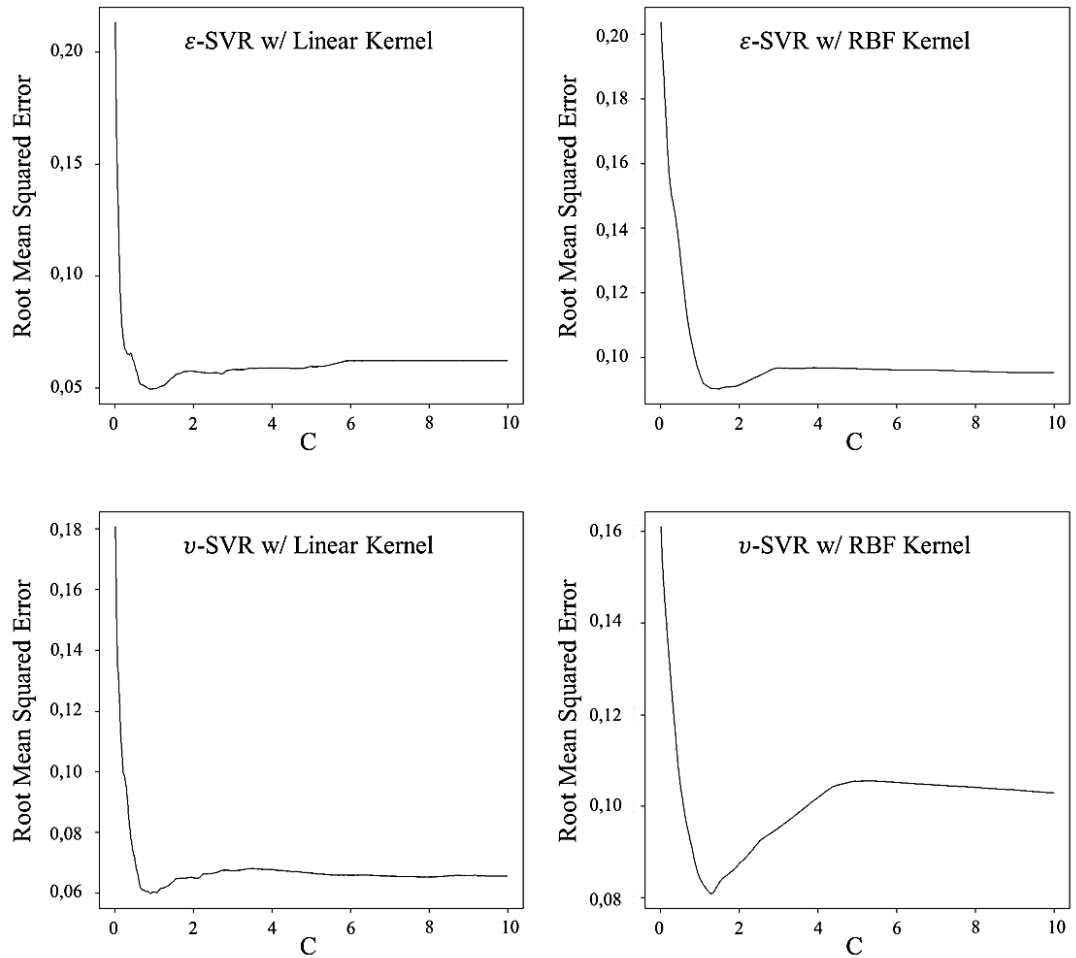


Figure 6.74 : Cost graphics of Case 5 for testing #4-7-10-18-22-27.

The basic statistical analysis of the whole data set is given in Table 6.85.

Table 6.85 : Statistics of the data set in Case 5 for testing #4-7-10-18-22-27.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,833	-2952	3135	-11916	-0,680	-0,248	-0,401	-0,213
Median	2,760	-2608	4160	-8342	-0,647	-0,261	-0,380	-0,095
Variance	0,199	6530787	7008892	66623679	0,174	0,035	0,060	0,181

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.86.

Table 6.86 : Statistical results of MR in Case 5 for learning and testing #4-7-10-18-22-27.

RSS _{default}	R ²	Adjusted R ²	p-value
1,14	0,89	0,85	1,11E-06

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.87.

Table 6.87 : Statistical results of the methods in Case 5 for testing #4-7-10-18-22-27.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,47	0,34	0,35	0,49	0,35
Pseudo-R ²	0,59	0,70	0,69	0,57	0,69
C		0,9	1,47	0,9	1,285
# of SV		20	21	16	19
CV Error		0,05	0,09	0,06	0,08
RMSE	0,28	0,24	0,24	0,29	0,24

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.88. The ROP values are in unit of [ft/hr].

Table 6.88 : ROP predictions in Case 5 for #4-7-10-18-22-27.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
4	17,8	16,8	16,2	18,3	15,8	10,0
7	16,4	16,1	17,6	16,8	17,6	13,0
10	17,9	17,4	17,5	18,0	17,8	15,7
18	39,6	37,1	34,1	35,6	32,7	38,6
22	24,3	21,3	18,8	23,1	18,9	19,0
27	13,0	13,2	13,2	13,2	13,5	12,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.75.

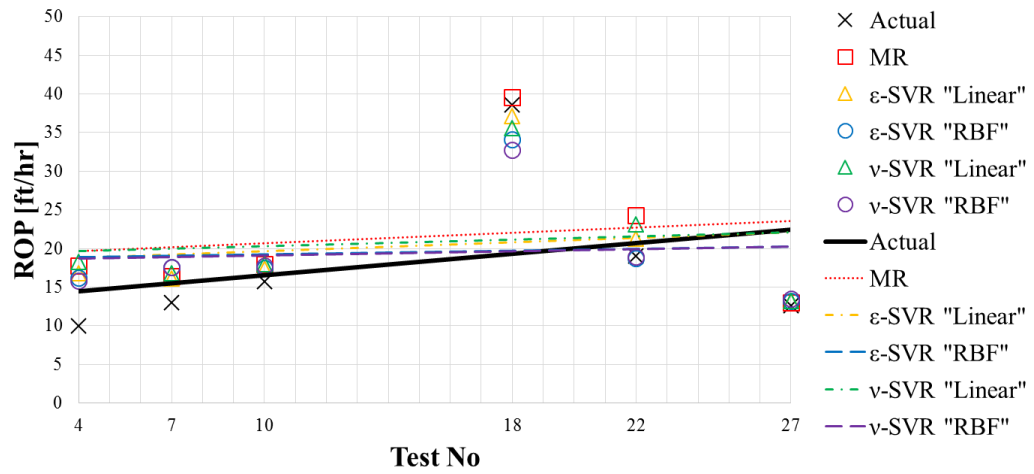


Figure 6.75 : ROP prediction trends in Case 5 for testing #4-7-10-18-22-27.

The comparison of each prediction with actual ROP is plotted in Figure 6.76.

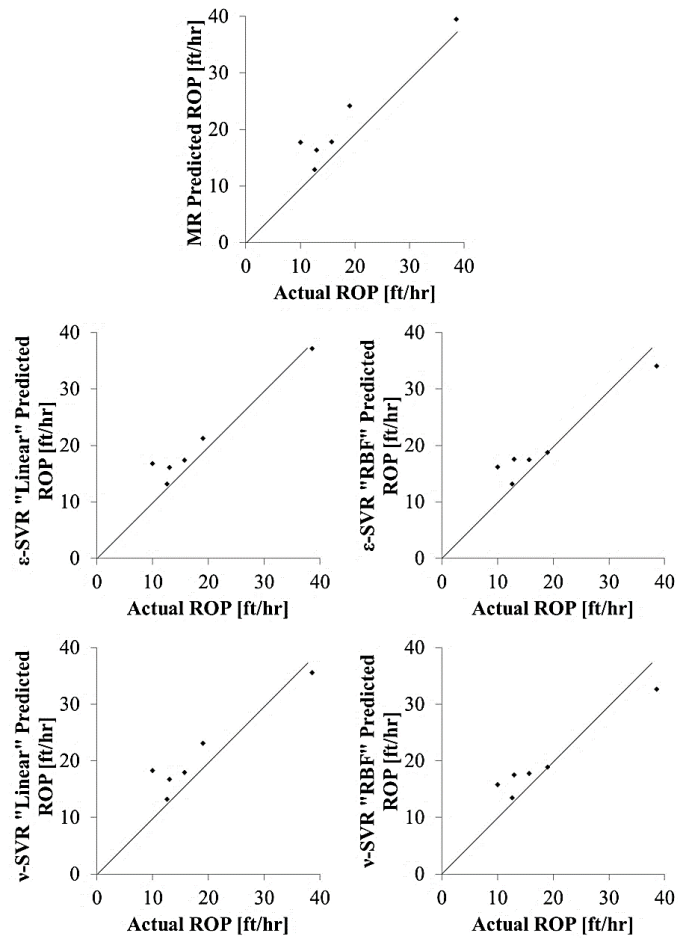


Figure 6.76 : ROP prediction comparison in Case 5 for testing #4-7-10-18-22-27.

For testing #4-7-10-18-22-27, ε -SVR with linear kernel and other RBF kernels can be considered as the best predictors. Similarly, these methods have all the minimum errors and maximum Pseudo- R^2 value.

The cumulative comparative statistical results for Case 5 are placed on several charts. Negative values are excluded from the graphics. The comparison of the scenarios in Case 4 in terms of RSS_{model} is shown in Figure 6.77.

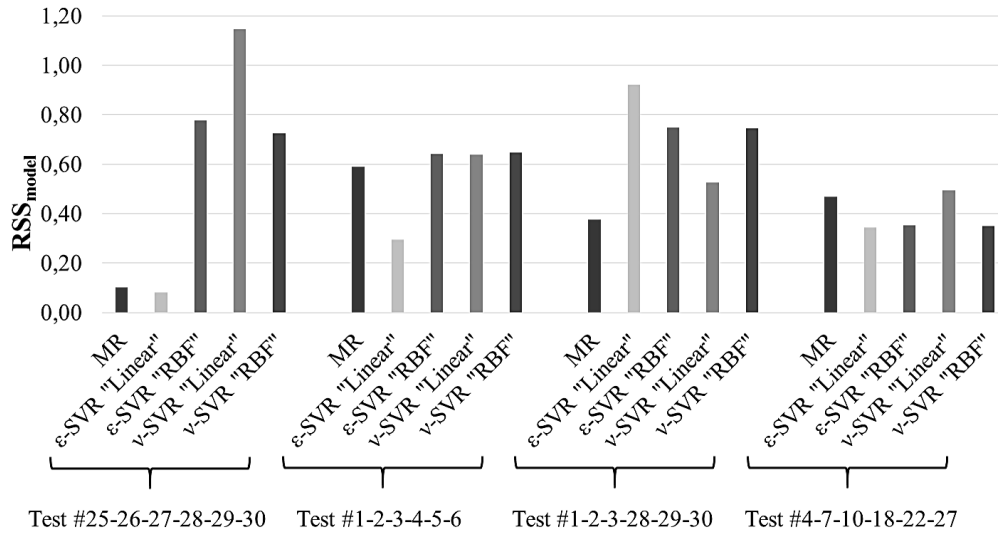


Figure 6.77 : RSS_{model} values for Case 5.

It can be seen from Figure 6.77 that ε -SVR with linear kernel has the minimum RSS values for most of the scenarios. On the other hand, RBF kernels and ν -SVR with linear kernel have maximum RSS in Case 5.

Additionally, for almost all the scenarios, MR gives the highest RSS value among other methods. This situation it can be said that, SVR methods with linear kernel may be considered as a better predictor than MR.

The comparison of the scenarios in Case 5 in terms of CV error is shown in Figure 6.78.

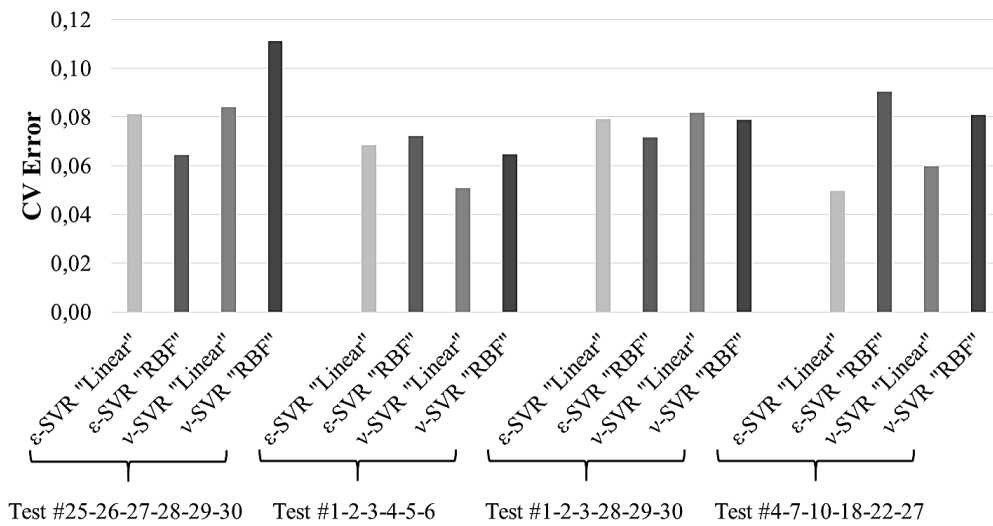


Figure 6.78 : CV error values for Case 5.

From Figure 6.78 it can be seen that, the results are variable, similar to the previous case. It is hard to make any generalization for this case in terms of CV error.

The testing errors of Case 5 in terms of RMSE is given in Figure 6.79.

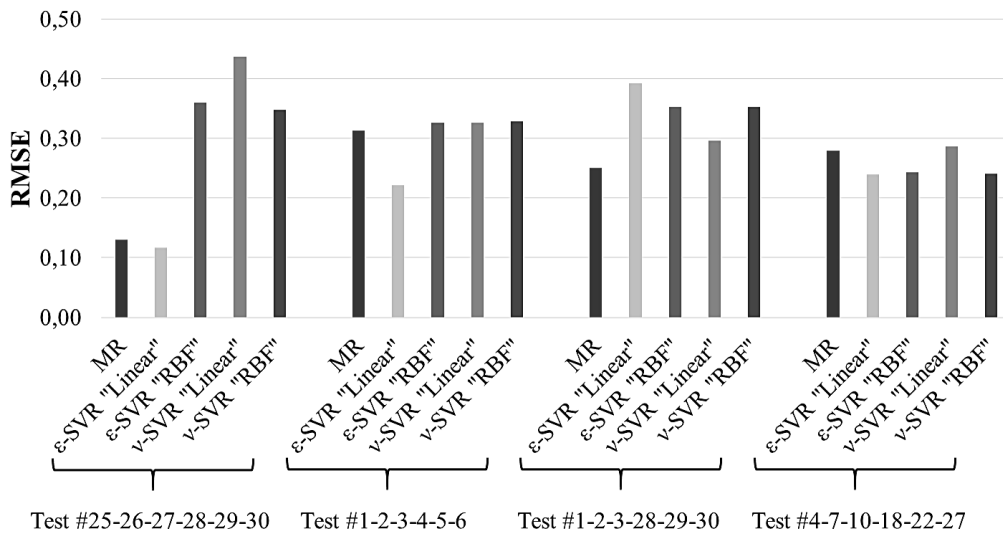


Figure 6.79 : Testing errors for Case 5.

As shown in Figure 6.79, the results are similar with Figure 6.77. ϵ -SVR with linear kernel can be taken into consideration as better predictor than other methods for most of the scenarios.

The Pseudo- R^2 comparison for Case 5 is given in Figure 6.80.

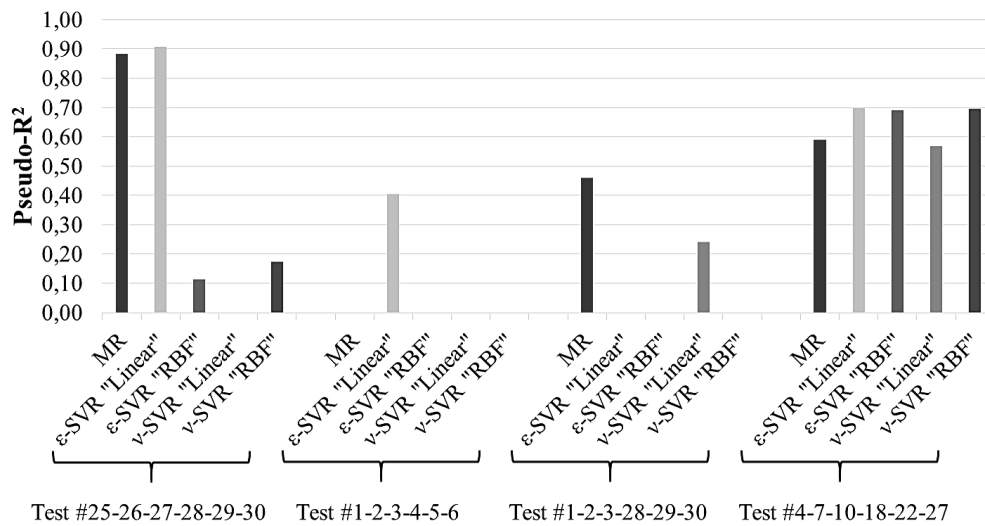


Figure 6.80 : Pseudo- R^2 values for Case 5.

It can be understood from Figure 6.80 that ϵ -SVR with linear kernel can be preferred as a predicting method for most of the scenarios. However, same as the previous cases, there are negative results, which means the models are not good enough to make accurate predictions. Similarly to the previous calculations, there are quite unstable results.

6.6 Case 6: Training All Data Points

In Case 6, all data points of Table A.1 are used as training data set. The comparison of the coefficients of MR analysis founded in this study with the study of Bourgoynne Jr. and Young Jr. (1974) is given in Table 6.89.

Table 6.89 : Multiple regression coefficients of Case 6.

	This Study	B&Y
a_1	3,77E-00	3,78E-00
a_2	0,18E-03	0,17E-03
a_3	0,20E-03	0,20E-03
a_4	0,43E-04	0,43E-04
a_5	0,42E-00	0,43E-00
a_6	0,18E-00	0,21E-00
a_7	0,41E-00	0,41E-00
a_8	0,16E-00	0,16E-00

There are slight differences between calculated coefficients and the results of the study of Bourgoynne Jr. and Young Jr. (1974), especially in the rotary speed exponent, a_6 .

6.6.1 Testing #1-2-3-4-5

In this testing scenario, the ROP values of the inputs numbered #1-2-3-4-5 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.81.

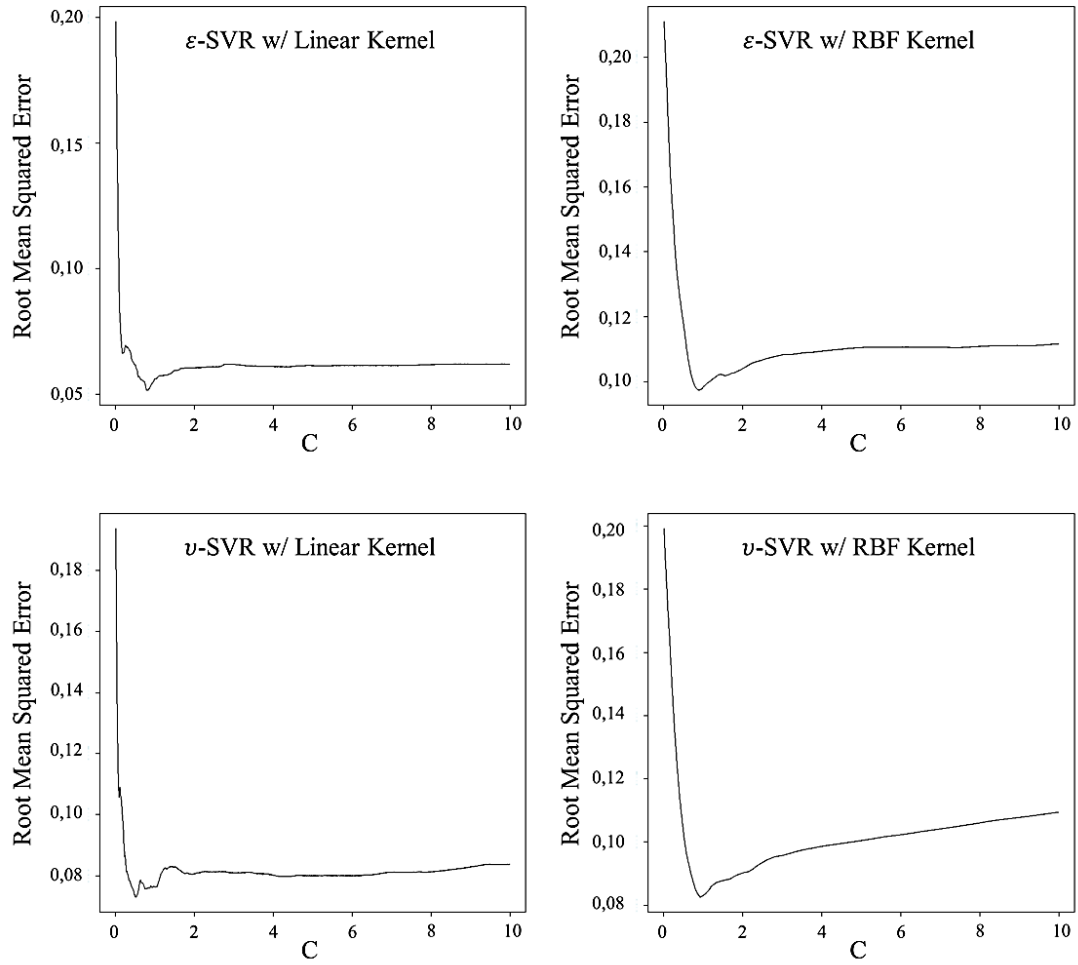


Figure 6.81 : Cost graphics of Case 6 for testing #1-2-3-4-5.

The basic statistical analysis of the whole data set is given in Table 6.90.

Table 6.90 : Statistics of the data set in Case 6 for testing #1-2-3-4-5.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,826	-2534	2687	-11077	-0,757	-0,206	-0,409	-0,191
Median	2,766	-1940	2211	-7833	-0,678	-0,249	-0,380	-0,037
Variance	0,184	6666268	7216663	61347257	0,209	0,046	0,061	0,158

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.91.

Table 6.91 : Statistical results of MR in Case 6 for learning and testing #1-2-3-4-5.

RSS_{default}	R^2	Adjusted R^2	p-value
0,48	0,86	0,81	5,47E-08

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.92.

Table 6.92 : Statistical results of the methods in Case 6 for testing #1-2-3-4-5.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS_{model}	0,22	0,23	0,17	0,23	0,19
Pseudo- R^2	0,54	0,51	0,64	0,53	0,60
C		0,83	0,9	0,505	0,915
# of SV		25	25	20	21
CV Error		0,05	0,10	0,07	0,08
RMSE	0,21	0,22	0,19	0,21	0,20

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.93. The ROP values are in unit of [ft/hr].

Table 6.93 : ROP predictions in Case 6 for #1-2-3-4-5.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
1	23,6	24,0	22,0	23,9	20,3	23,0
2	19,0	20,7	18,7	20,4	19,1	22,0
3	12,5	13,4	14,6	13,5	15,9	14,0
4	15,2	16,0	14,4	15,9	14,2	10,0
5	14,4	14,8	14,4	14,9	14,1	16,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.82.

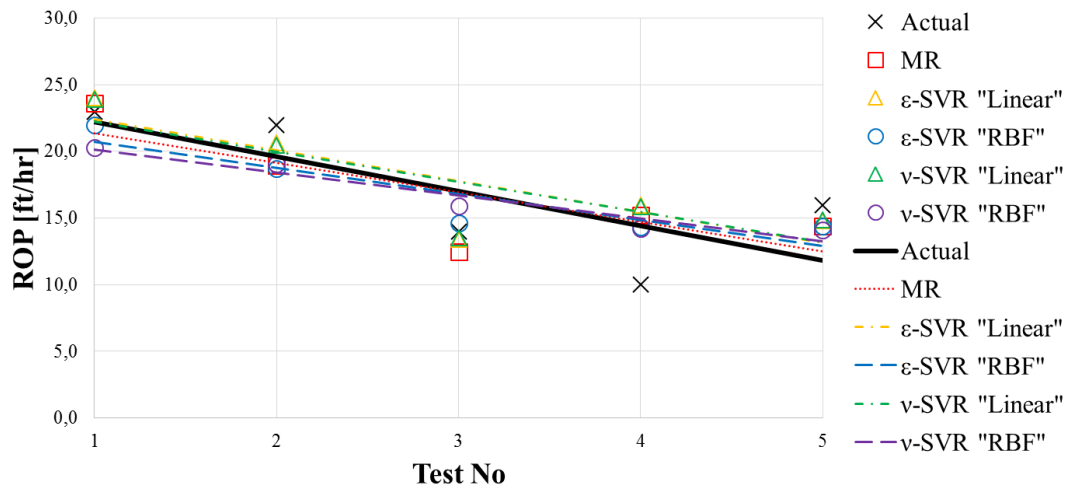


Figure 6.82 : ROP prediction trends in Case 6 for testing #1-2-3-4-5.

The comparison of each prediction with actual ROP is plotted in Figure 6.83.

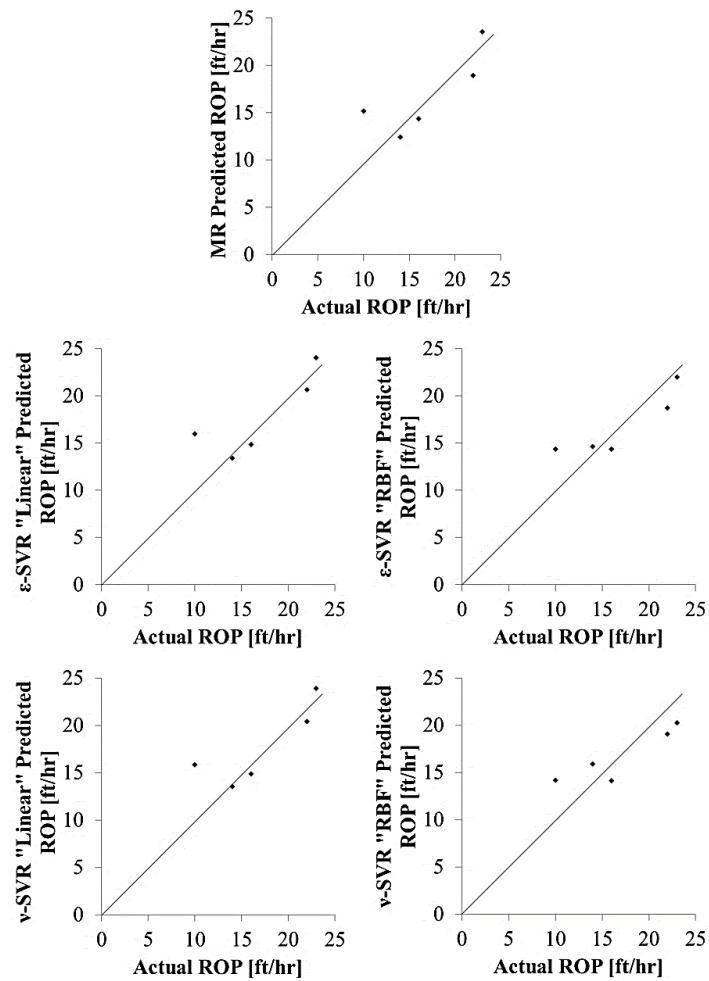


Figure 6.83 : ROP prediction comparison in Case 6 for testing #1-2-3-4-5.

It can be seen from Figure 6.82 and 6.83 that the best predictive methods in Case 6 for testing #1-2-3-4-5 are ε -SVR with linear and RBF kernels, since they have the best correlation between predicted and actual ROP values. Additionally, ε -SVR with RBF kernel has the minimum errors and maximum Pseudo- R^2 ratio among the other methods, as seen in the results.

6.6.2 Testing #26-27-28-29-30

In this testing scenario, the ROP values of the inputs numbered #26-27-28-29-30 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.84.

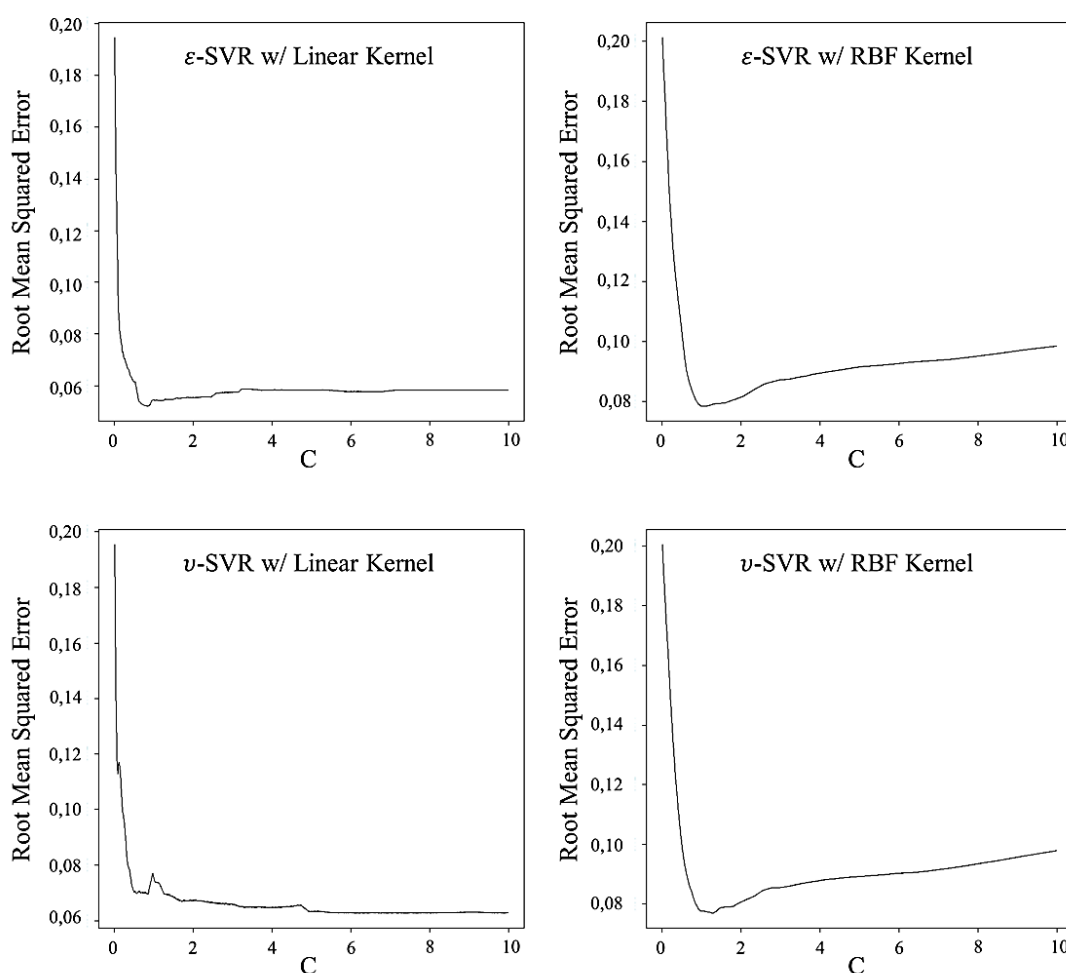


Figure 6.84 : Cost graphics of Case 6 for testing #26-27-28-29-30.

The basic statistical analysis of the whole data set is given in Table 6.94.

Table 6.94 : Statistics of the data set in Case 6 for testing #26-27-28-29-30.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,792	-3546	3586	-12764	-0,620	-0,259	-0,411	-0,246
Median	2,701	-3055	4980	-10118	-0,549	-0,261	-0,380	-0,130
Variance	0,185	8141601	7266751	63638495	0,177	0,032	0,071	0,166

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.95.

Table 6.95 : Statistical results of MR in Case 6 for learning and testing #26-27-28-29-30.

RSS _{default}	R ²	Adjusted R ²	p-value
0,57	0,86	0,81	5,47E-08

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.96.

Table 6.96 : Statistical results of the methods in Case 6 for testing #26-27-28-29-30.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,05	0,04	0,01	0,06	0,04
Pseudo-R ²	0,92	0,93	0,97	0,89	0,94
C		0,83	0,995	6,185	1,27
# of SV		25	25	19	21
CV Error		0,05	0,08	0,06	0,08
RMSE	0,10	0,09	0,05	0,11	0,09

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.97. The ROP values are in unit of [ft/hr].

Table 6.97 : ROP predictions in Case 6 for #26-27-28-29-30.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
26	15,8	15,5	15,5	16,0	16,2	14,8
27	12,9	13,2	13,2	13,4	13,8	12,6
28	17,7	17,2	15,6	18,2	16,3	14,9
29	12,6	12,2	14,1	12,7	14,6	13,8
30	8,6	9,4	9,8	9,7	9,9	9,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.85.

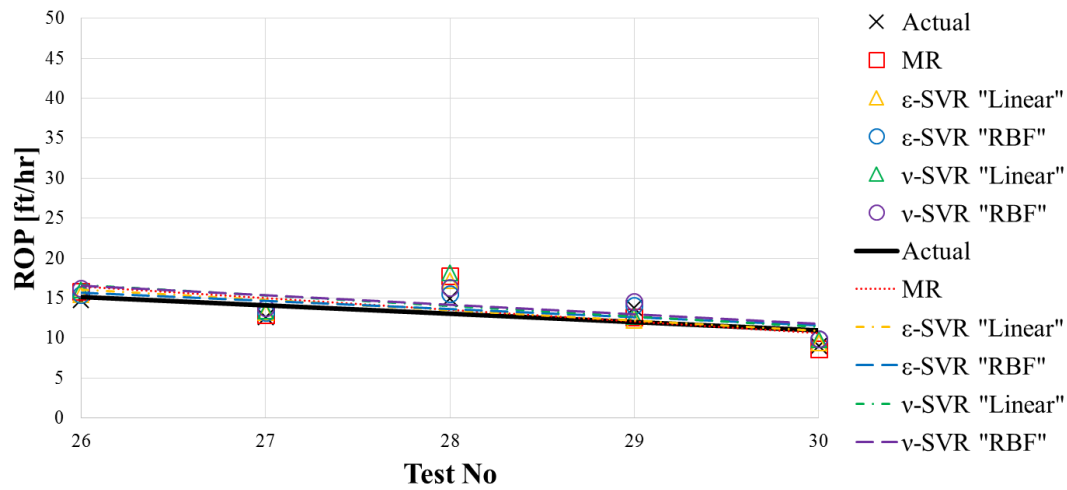


Figure 6.85 : ROP prediction trends in Case 6 for testing #26-27-28-29-30.

The comparison of each prediction with actual ROP is plotted in Figure 6.86.

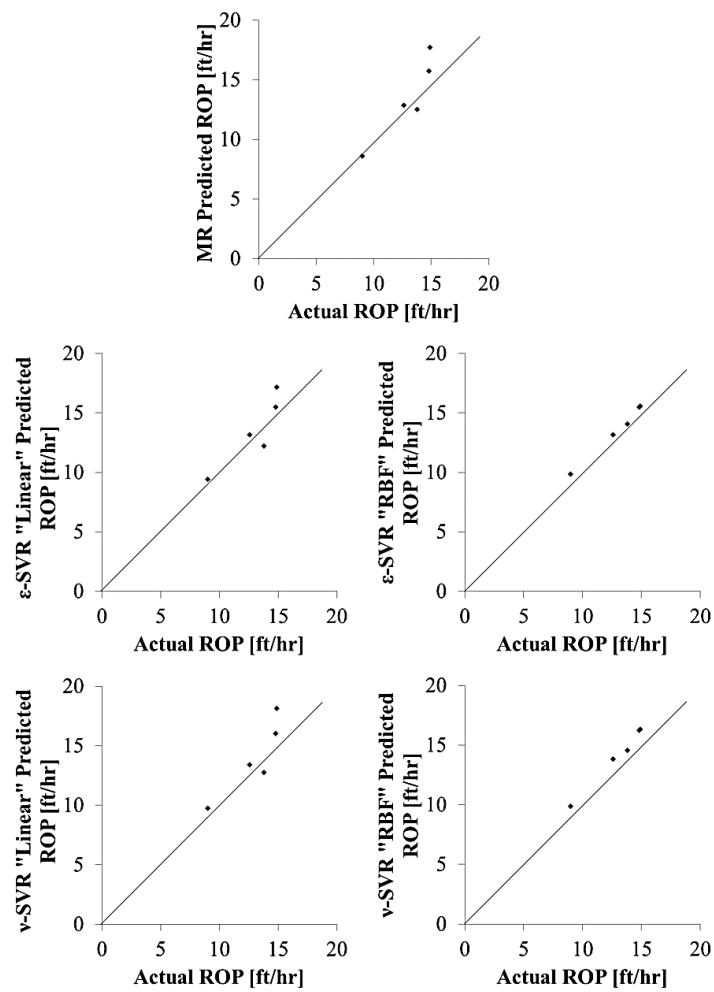


Figure 6.86 : ROP prediction comparison in Case 6 for testing #26-27-28-29-30.

As seen on Figure 6.85 and 6.86, the highest correlation is in ε -SVR with RBF kernel. Moreover, ε -SVR with RBF kernel has minimum RMSE and RSS, plus maximum Pseudo- R^2 value. Additionally, it is noticeable that the performance of MR increased since training data number is set to 30.

6.6.3 Testing #14-15-16-17-18

In this testing scenario, the ROP values of the inputs numbered #14-15-16-17-18 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.87.

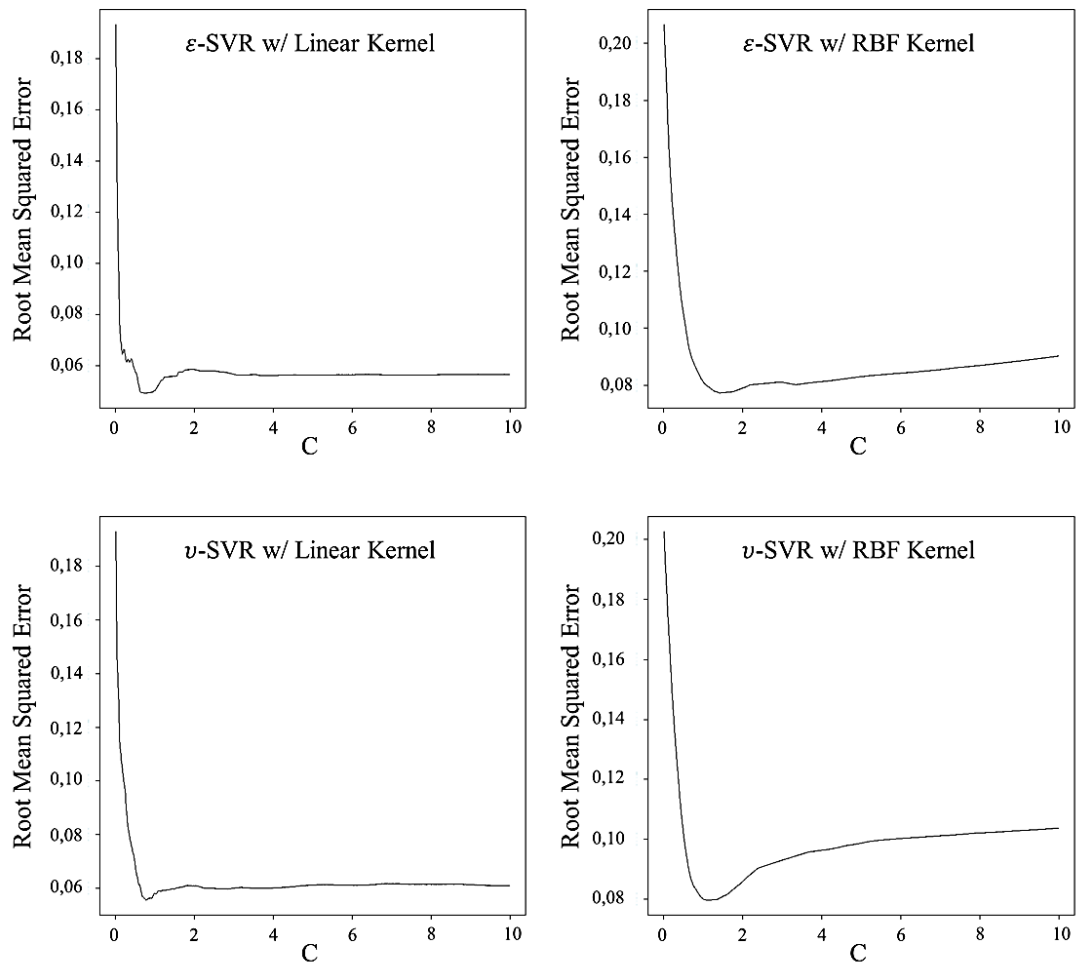


Figure 6.87 : Cost graphics of Case 6 for testing #14-15-16-17-18.

The basic statistical analysis of the whole data set is given in Table 6.98.

Table 6.98 : Statistics of the data set in Case 6 for testing #14-15-16-17-18.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,881	-2911	3281	-12555	-0,629	-0,248	-0,383	-0,164
Median	2,766	-2900	4733	-8406	-0,571	-0,261	-0,380	-0,037
Variance	0,233	5603844	6235732	73921080	0,170	0,030	0,056	0,170

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.99.

Table 6.99 : Statistical results of MR in Case 6 for learning and testing #14-15-16-17-18.

RSS _{default}	R ²	Adjusted R ²	p-value
2,23	0,86	0,81	5,47E-08

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.100.

Table 6.100 : Statistical results of the methods in Case 6 for testing #14-15-16-17-18.

	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,01	0,01	0,03	0,02	0,08
Pseudo-R ²	0,997	0,996	0,987	0,990	0,964
C		0,785	1,415	0,775	1,12
# of SV		25	23	20	21
CV Error		0,05	0,08	0,06	0,08
RMSE	0,04	0,04	0,08	0,07	0,13

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.101. The ROP values are in unit of [ft/hr].

Table 6.101 : ROP predictions in Case 6 for #14-15-16-17-18.

Test No	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"	Actual
14	8,9	9,2	9,9	8,9	10,7	9,6
15	15,8	15,2	16,2	14,9	16,5	15,5
16	31,0	30,1	31,7	29,5	30,4	31,4
17	42,2	40,8	36,6	39,5	34,0	42,7
18	38,8	36,9	36,9	36,1	34,6	38,6

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.88.

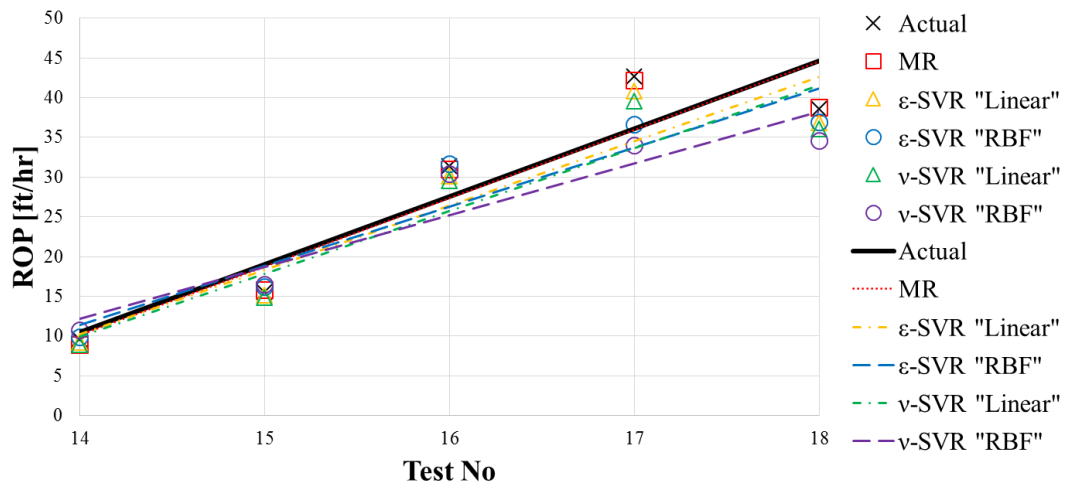


Figure 6.88 : ROP prediction trends in Case 6 for testing #14-15-16-17-18.

The comparison of each prediction with actual ROP is plotted in Figure 6.89.

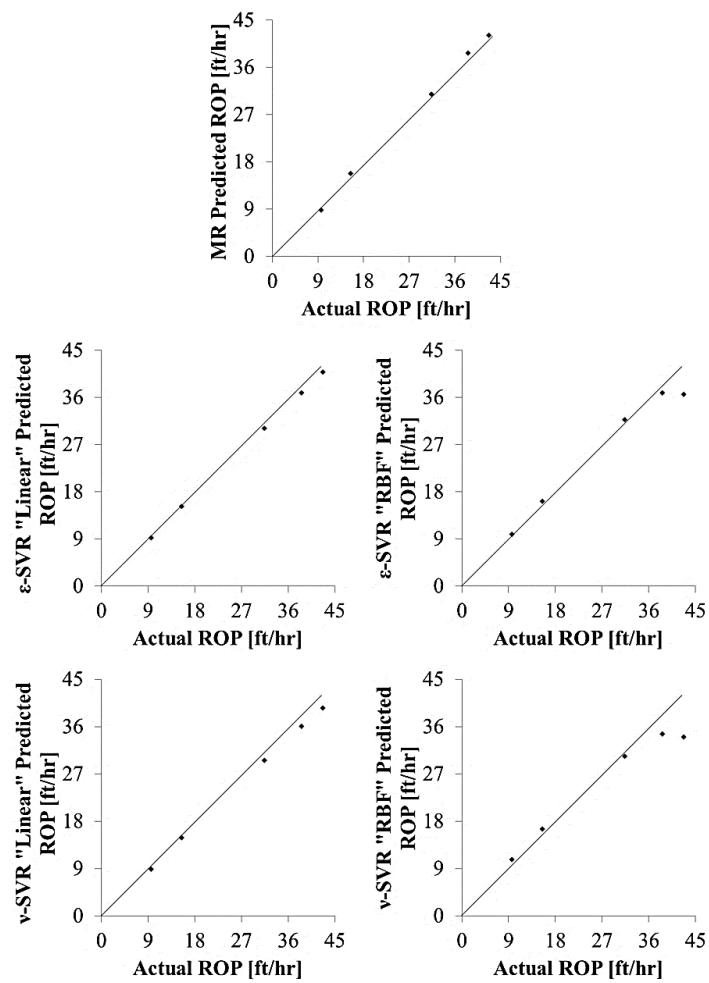


Figure 6.89 : ROP prediction comparison in Case 6 for testing #14-15-16-17-18.

According to the Figures 6.88 and 6.89, the best predictor method for this scenario is MR. Furthermore, the minimum RMSE and RSS with maximum Pseudo- R^2 value are in MR. Additionally, the second best results are in ε -SVR with linear kernel. For this scenario, MR and ε -SVR with linear kernel give similar accuracy in terms of ROP prediction. However, the best method is different than the previous scenario.

6.6.4 Testing #1-6-11-16-21

In this testing scenario, the ROP values of the inputs numbered #1-6-11-16-21 are desired to be predicted.

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.90.

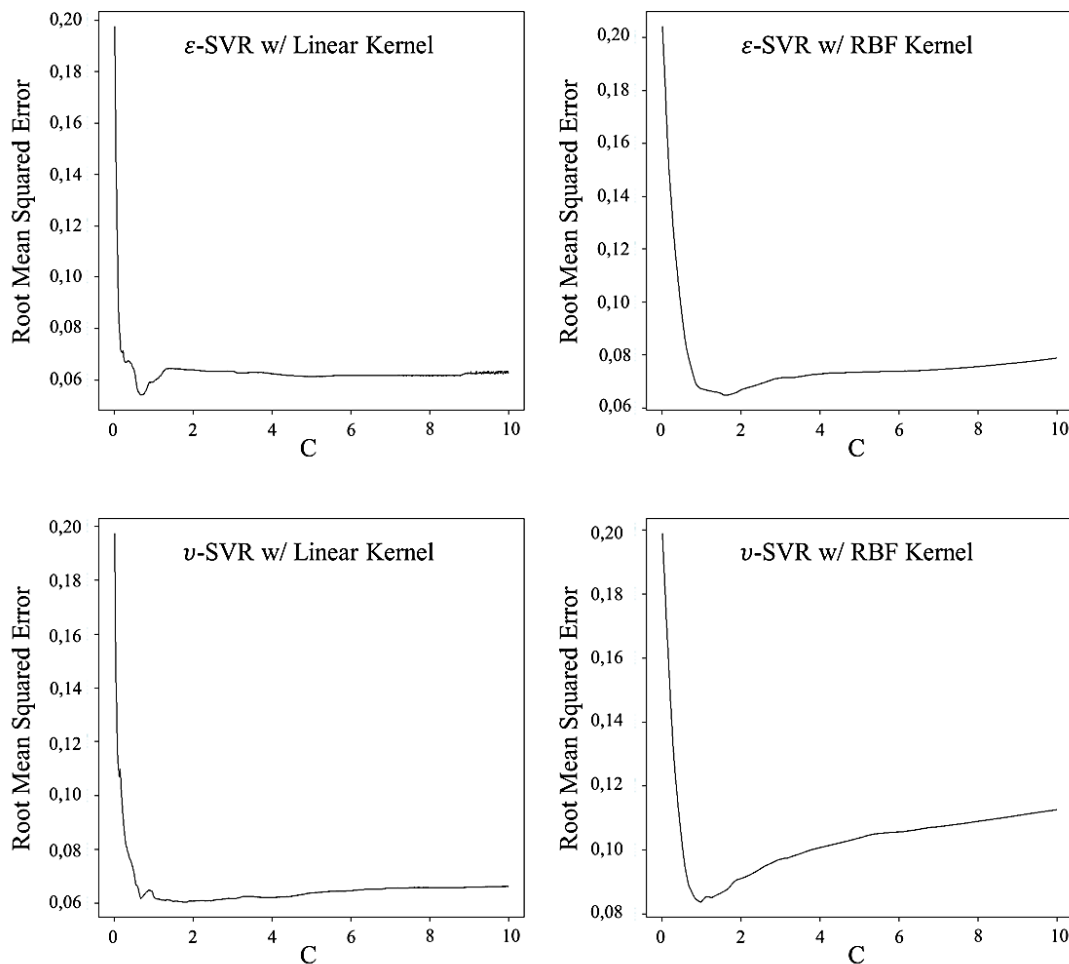


Figure 6.90 : Cost graphics of Case 6 for testing #1-6-11-16-21.

The basic statistical analysis of the whole data set is given in Table 6.102.

Table 6.102 : Statistics of the data set in Case 6 for testing #1-6-11-16-21.

	y	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Mean	2,863	-2770	2982	-11327	-0,694	-0,235	-0,414	-0,229
Median	2,773	-2315	3588	-8406	-0,659	-0,261	-0,400	-0,070
Variance	0,186	6157823	6959063	59561363	0,169	0,034	0,058	0,217

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.103.

Table 6.103 : Statistical results of MR in Case 6 for learning and testing #1-6-11-16-21.

RSS _{default}	R ²	Adjusted R ²	p-value
0,58	0,86	0,81	5,47E-08

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.104.

Table 6.104 : Statistical results of the methods in Case 6 for testing #1-6-11-16-21.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	0,17	0,42	0,02	0,23	0,07
Pseudo-R ²	0,70	0,26	0,97	0,59	0,88
C		24	1,655	1,805	0,98
# of SV		20	21	19	21
CV Error		0,05	0,06	0,06	0,08
RMSE	0,18	0,29	0,06	0,22	0,12

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.105. The ROP values are in unit of [ft/hr].

Table 6.105 : ROP predictions in Case 6 for #1-6-11-16-21.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
1	23,6	24,1	22,0	24,2	20,4	10,0
6	15,9	16,4	17,3	16,3	16,8	13,0
11	14,4	15,4	15,1	15,2	15,8	15,7
16	31,0	30,0	31,8	29,3	30,1	38,6
21	14,6	11,4	20,2	13,5	18,3	19,0

The cumulative comparison of actual and predicted ROP values is plotted on Figure 6.91.

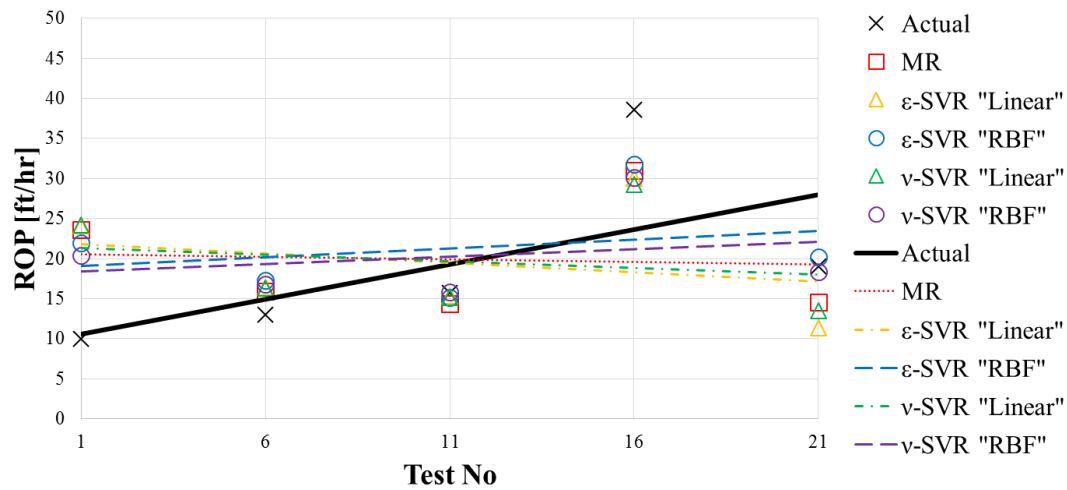


Figure 6.91 : ROP prediction trends in Case 6 for testing #1-6-11-16-21.

The comparison of each prediction with actual ROP is plotted in Figure 6.92.

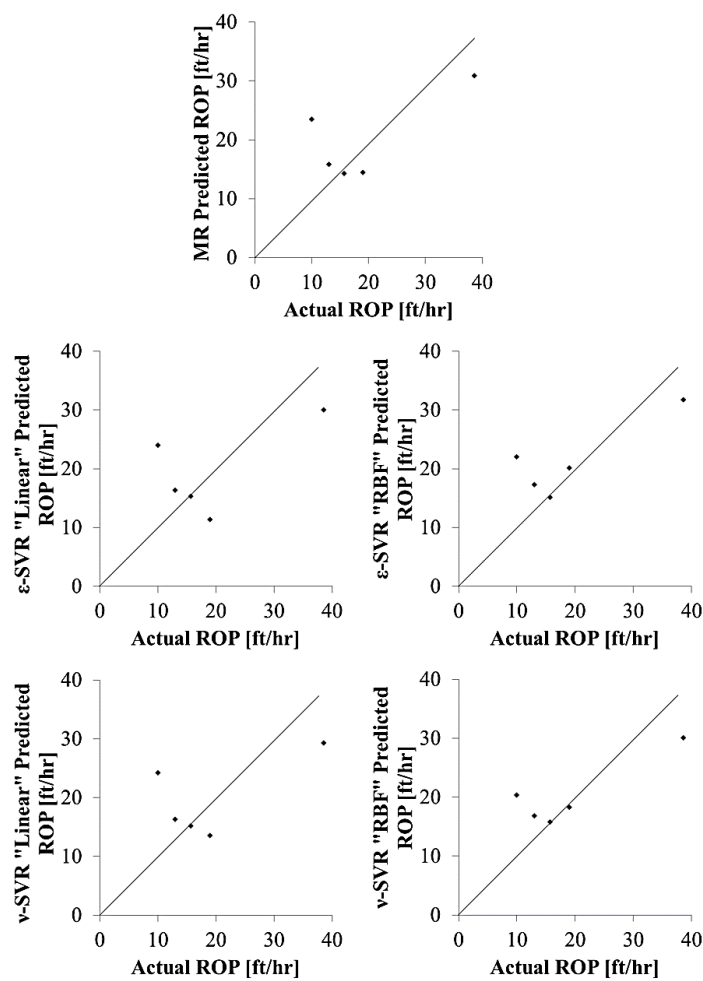


Figure 6.92 : ROP prediction comparison in Case 6 for testing #1-6-11-16-21.

Correspondingly to previous scenarios, ε -SVR with RBF kernel is the best predictor. Likewise, ε -SVR with RBF kernel has all the minimum errors and maximum Pseudo- R^2 value. In addition, the best predictors are seemed to be RBF kernel methods rather than MR for this scenario.

It should be noticed that the most accurate methods in prediction changes in every different scenario, although the training set remains same in Case 6. Similar to the previous cases, an appropriate generalization is not achievable due to varying indicators.

The cumulative comparative statistical results for Case 6 are placed on several charts. Negative values are excluded from the graphics. The comparison of the scenarios in Case 6 in terms of RSS_{model} is shown in Figure 6.93.

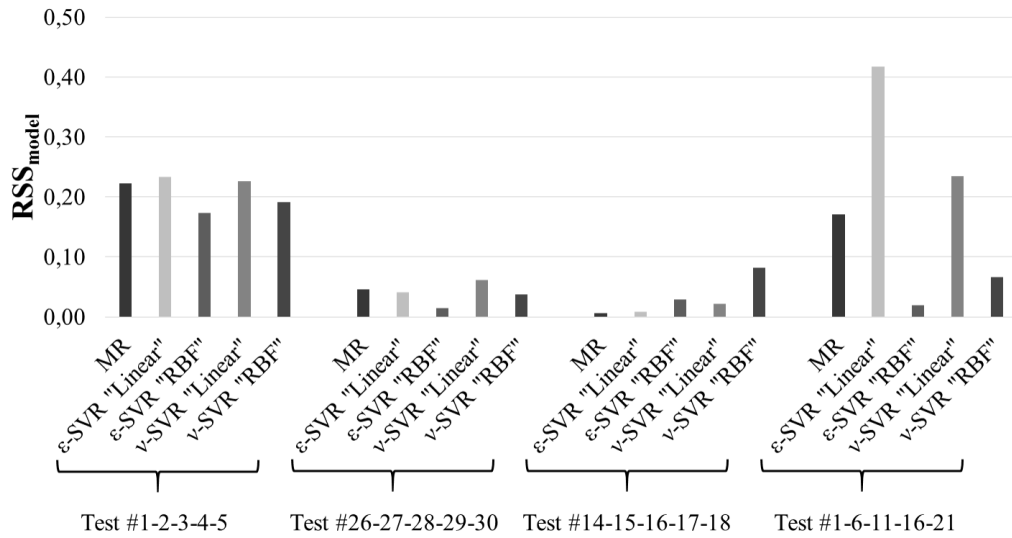


Figure 6.93 : RSS_{model} values for Case 6.

It can be seen from Figure 6.93 that RSS values are variable. The results are data dependent, and it is not a good approximation to make a generalization on the RSS of methods.

The comparison of the scenarios in Case 6 in terms of CV error is shown in Figure 6.94.

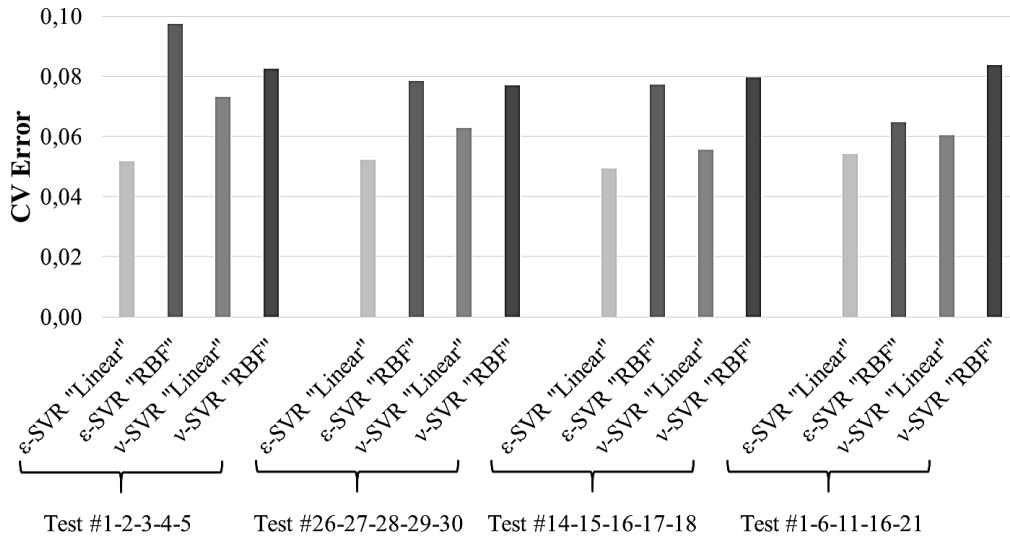


Figure 6.94 : CV error values for Case 6.

In overall, SVR methods with linear kernel have the best cross validation results.

The testing errors of Case 6 in terms of RMSE is given in Figure 6.95.

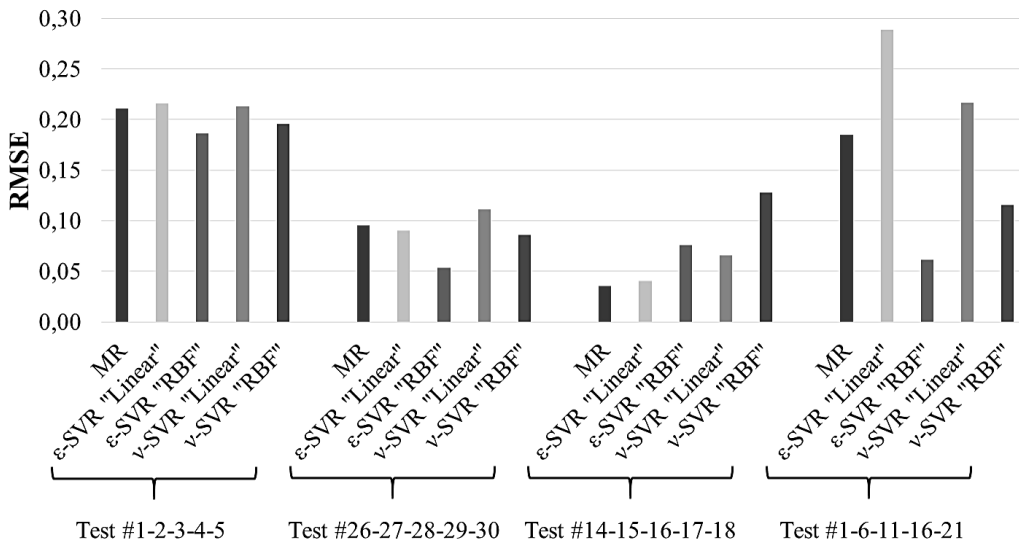


Figure 6.95 : Testing errors for Case 6.

As shown in Figure 6.95, SVR methods with RBF kernel have the minimum testing error in most of the scenarios.

The Pseudo- R^2 comparison for Case 6 is given in Figure 6.96.

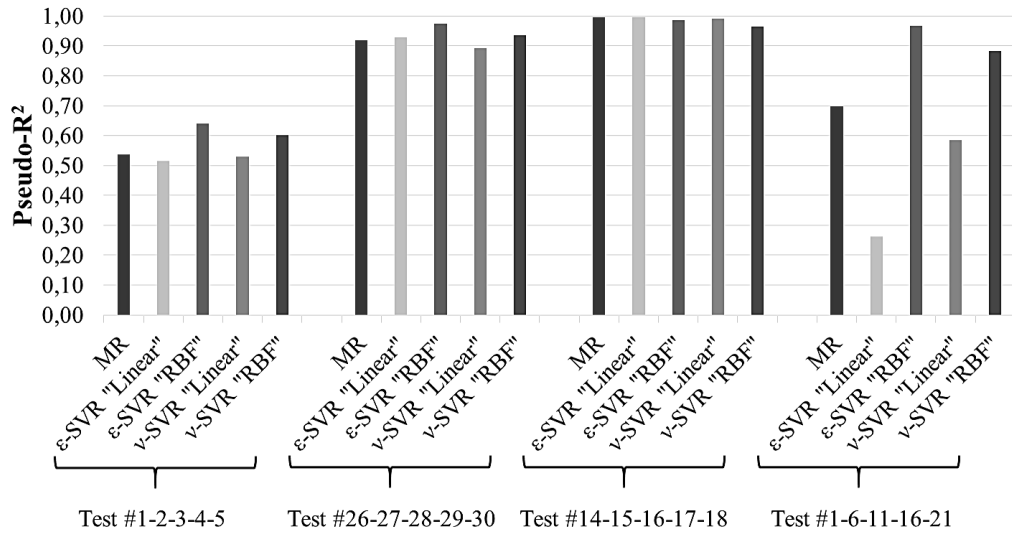


Figure 6.96 : Pseudo- R^2 values for Case 6.

It can be understood from Figure 6.96 that SVR with RBF shows better accuracy in predicting testing data for Case 6. Furthermore, it can be understood that, the accuracy of prediction may vary with changing test dataset.

6.6.5 Testing all

In this testing scenario, all the data in Table A.1 is desired to be predicted.

The basic statistical analysis of the whole data set is given in Table 6.106.

Table 6.106 : Statistics of the data set in Case 6 for testing all.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	2,833	-2952	3135	-11916	-0,680	-0,248	-0,401	-0,213
Median	2,760	-2608	4160	-8342	-0,647	-0,261	-0,380	-0,095
Variance	0,196	6420096	6890097	65494464	0,171	0,034	0,059	0,178

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.107.

Table 6.107 : Statistical results of MR in Case 6 for learning and testing all.

RSS _{default}	R ²	Adjusted R ²	p-value
5,78	0,86	0,81	5,47E-8

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.97.

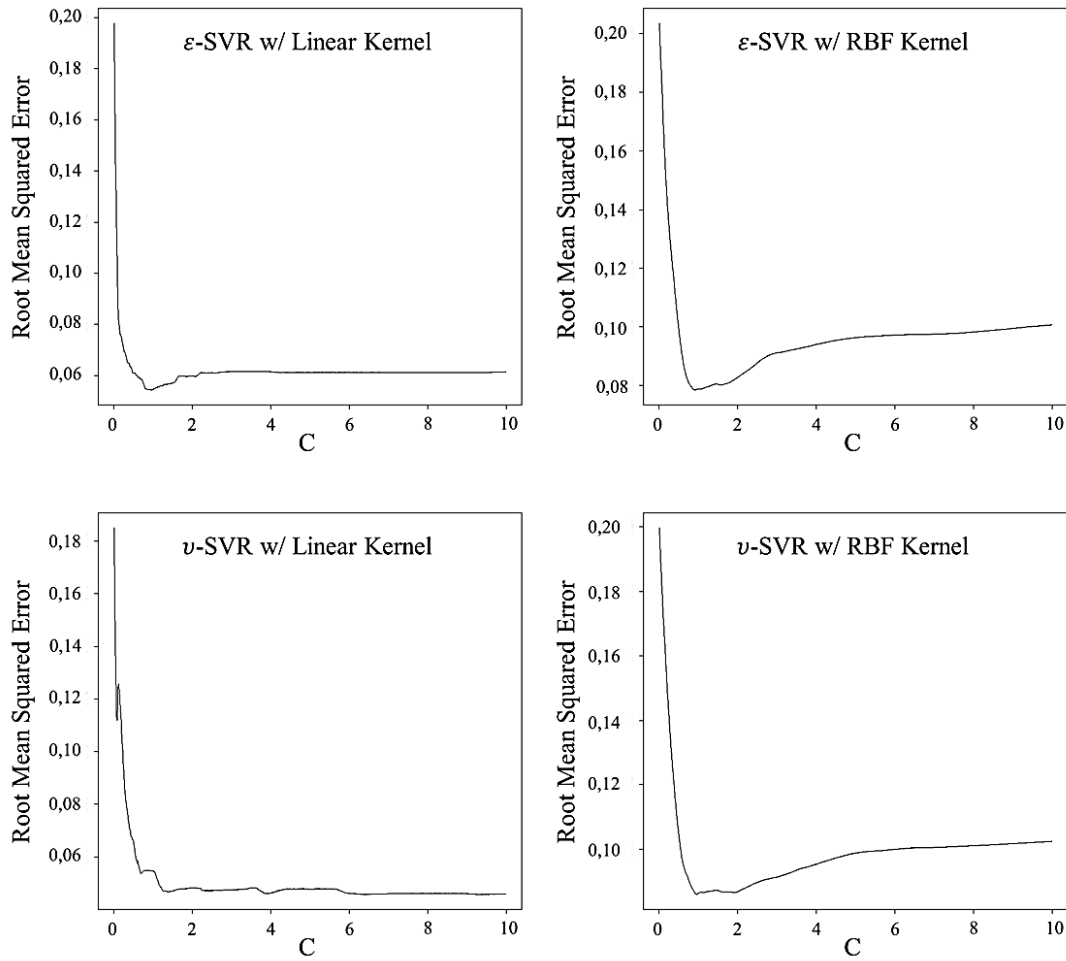


Figure 6.97 : Cost graphics of Case 6 for testing all.

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.108.

Table 6.108 : Statistical results of the methods in Case 6 for testing all.

	MR	ϵ -SVR "Linear"	ϵ -SVR "RBF"	ν -SVR "Linear"	ν -SVR "RBF"
RSS _{model}	0,82	0,97	0,73	0,92	0,64
Pseudo-R ²	0,86	0,83	0,87	0,84	0,89
C		0,995	0,91	0,75	1,295
# of SV		25	25	20	21
CV Error		0,06	0,08	0,06	0,09
RMSE	0,16	0,18	0,16	0,17	0,15

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.109. The ROP values are in unit of [ft/hr].

Table 6.109 : ROP predictions in Case 6 for all.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
1	23,6	24,1	22,0	24,4	21,0	23,0
2	19,0	20,7	18,7	20,3	20,1	22,0
3	12,5	13,4	14,6	12,9	15,3	14,0
4	15,2	16,0	14,3	15,8	14,0	10,0
5	14,4	14,8	14,4	14,8	14,0	16,0
6	15,9	16,1	17,5	16,2	17,4	19,0
7	15,3	15,4	16,5	15,5	16,4	13,0
8	16,5	16,4	16,3	16,8	16,2	16,6
9	16,4	15,9	16,2	16,4	16,7	15,9
10	16,9	16,5	16,4	17,0	16,9	15,7
11	14,4	15,1	16,0	15,2	15,3	14,0
12	10,3	10,2	12,9	10,3	12,3	13,5
13	7,3	7,2	9,3	7,0	8,8	6,2
14	8,9	9,2	10,0	8,9	10,2	9,6
15	15,8	15,2	16,2	15,5	16,0	15,5
16	31,0	30,1	30,2	31,7	30,8	31,4
17	42,2	40,8	34,1	44,0	34,7	42,7
18	38,8	36,9	34,4	39,7	35,3	38,6
19	33,4	31,9	31,6	34,1	32,6	43,4
20	18,4	15,1	15,4	16,5	16,3	12,5
21	14,6	11,5	17,6	12,7	19,3	21,1
22	21,7	20,2	18,4	21,6	19,2	19,0
23	19,5	19,1	19,6	19,9	20,5	18,7
24	19,3	17,2	16,7	18,6	18,2	20,2
25	26,4	28,3	25,9	29,3	26,0	27,1
26	15,8	15,5	15,5	16,0	16,2	14,8
27	12,9	13,2	13,2	13,4	13,8	12,6
28	17,7	17,2	15,6	18,2	16,3	14,9
29	12,6	12,2	14,1	12,7	14,5	13,8
30	8,6	9,4	10,1	9,7	9,8	9,0

The cumulative comparison of actual and predicted ROP values versus depth is plotted on Figure 6.98.

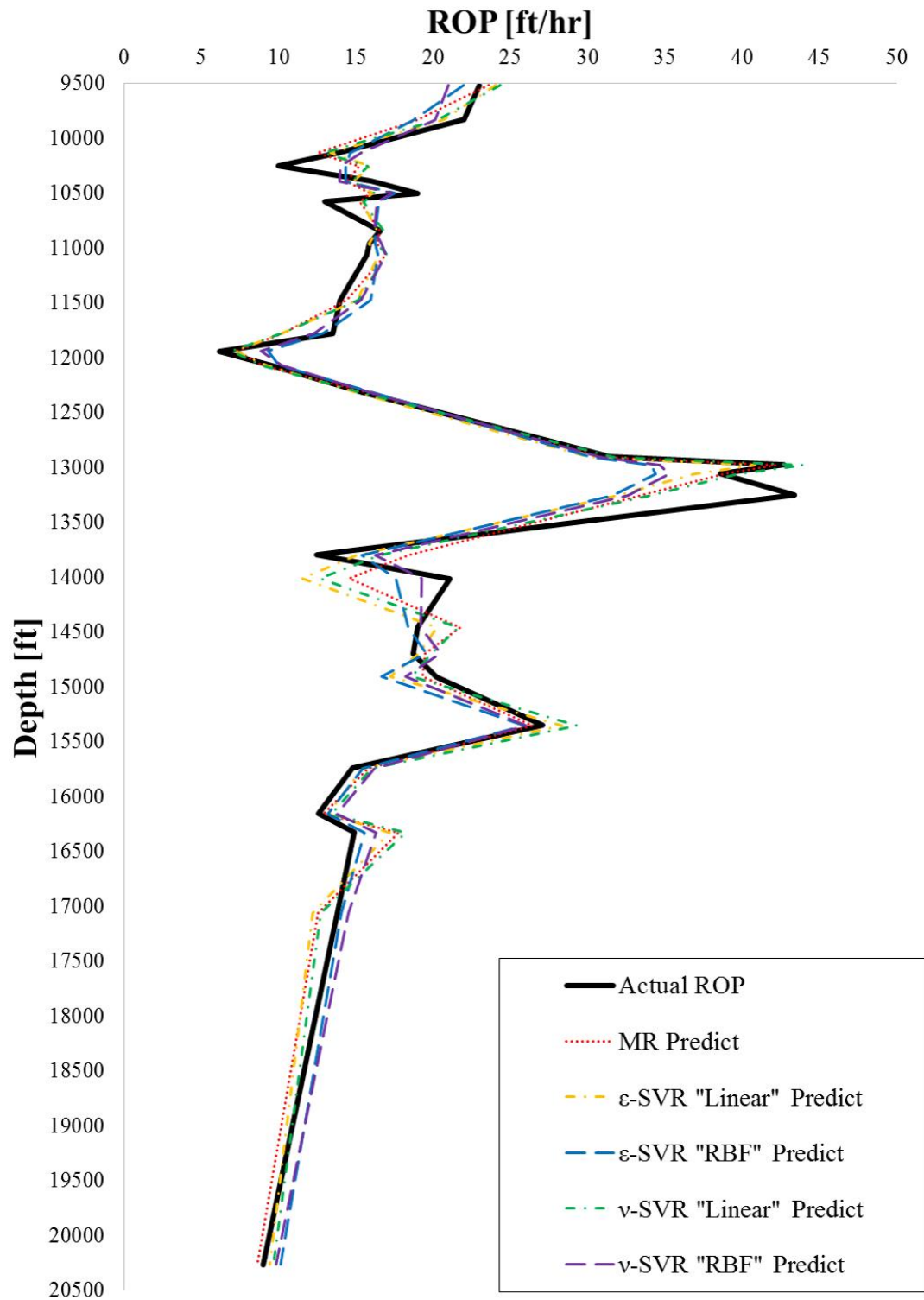


Figure 6.98 : ROP prediction trends in Case 6 for testing all.

The comparison of each prediction with actual ROP is plotted in Figure 6.99.

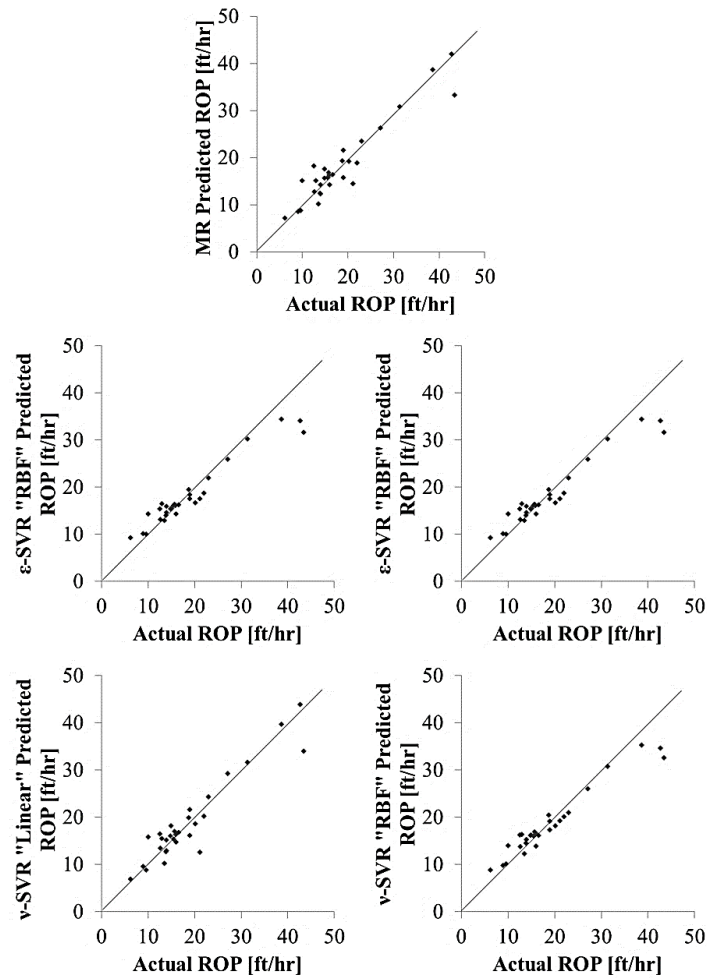


Figure 6.99 : ROP prediction comparison in Case 6 for testing all.

It can be understood from Figures 6.98 and 6.99 that, ν -SVR with RBF kernel seemed to be the best predictor for the whole data set, since it has the minimum RMSE, RSS and maximum Pseudo- R^2 .

6.7 Case 7: Training Different Data Set

In this case, all the data in Table A.2 is desired to be predicted.

The basic statistical analysis of the whole data set is given in Table 6.110.

Table 6.110 : Statistics of the data set in Case 7 for testing all.

	y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
Mean	3,351	1362	-4	-9635	-0,890	0,217	-0,335	0,009
Median	3,332	637	-127	-9727	-0,821	0,182	-0,250	0,060
Variance	0,880	11954156	120896	22371525	0,285	0,087	0,123	0,049

The coefficients of MR analysis for this training data are given in Table 6.111.

Table 6.111 : Multiple regression coefficients in Case 7 for testing all.

a1	a2	a3	a4	a5	a6	a7	a8
3,907E+00	9,453E-05	6,862E-05	8,642E-05	3,731E-01	2,232E+00	2,480E-02	6,701E-01

The statistical results of the multiple regression analysis for the given testing data set are expressed in Table 6.112.

Table 6.112 : Statistical results of MR in Case 7 for learning and testing all.

RSS_{default}	R^2	Adjusted R^2	p-value
21,57	0,72	0,60	0,001

The plots of the determination of cost parameter via 10-fold cross validation are shown in Figure 6.100.

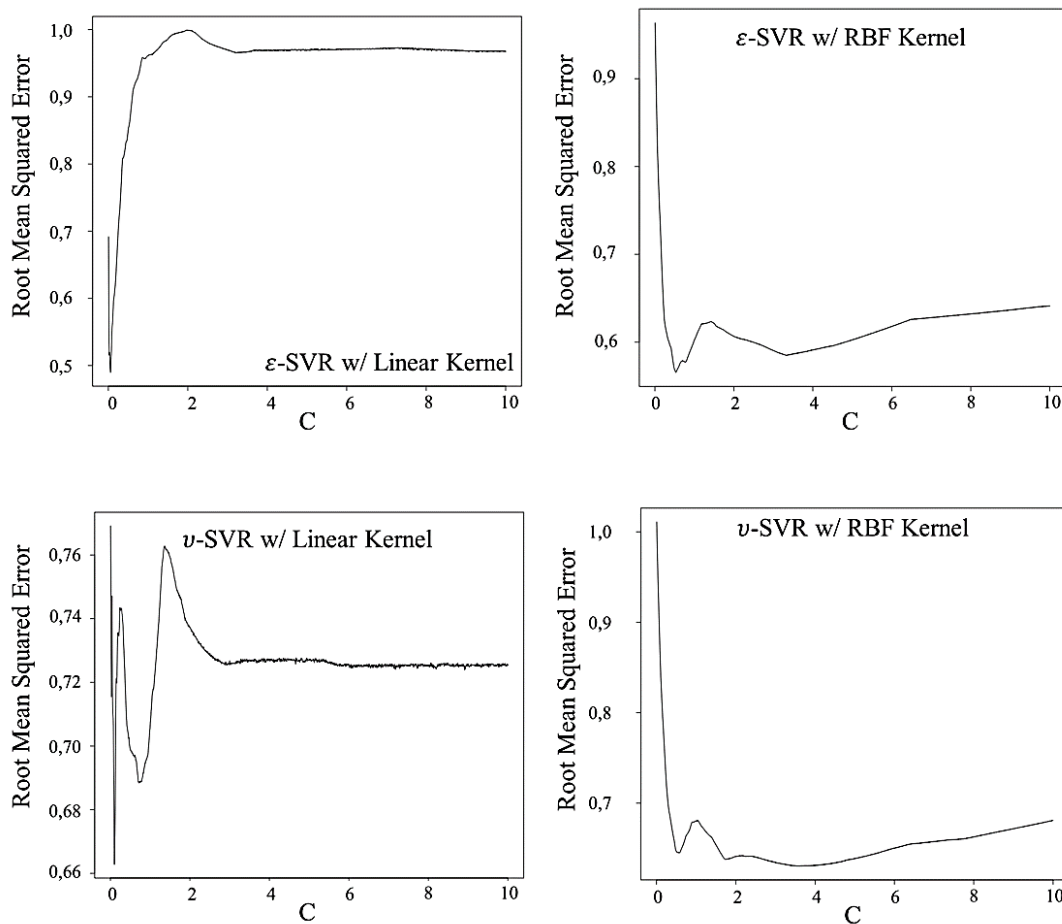


Figure 6.100 : Cost graphics of Case 7 for testing all.

The statistical results of different predictive methods for the given testing data set by using current training data set are represented in Table 6.113.

Table 6.113 : Statistical results of the methods in Case 7 for testing all.

	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”
RSS _{model}	6,07	9,17	7,53	8,26	2,18
Pseudo-R ²	0,72	0,57	0,65	0,62	0,90
C		0,055	0,53	0,1	3,61
# of SV		22	21	16	22
CV Error		0,49	0,57	0,66	0,63
RMSE	0,49	0,61	0,55	0,57	0,30

The actual and predicted ROP values of the testing data set by using current training set and different predictive methods are given in Table 6.113. The ROP values are in unit of [ft/hr].

Table 6.114 : ROP predictions in Case 7 for all.

Test No	MR	ϵ -SVR “Linear”	ϵ -SVR “RBF”	ν -SVR “Linear”	ν -SVR “RBF”	Actual
1	70	64	48	74	123	171
2	47	61	29	64	22	20
3	70	63	59	75	148	160
4	69	50	59	63	89	82
5	72	45	49	58	47	49
6	71	47	47	56	47	43
7	43	41	58	47	69	64
8	35	33	34	37	33	36
9	35	31	30	35	28	27
10	24	47	27	49	15	14
11	78	55	58	64	77	83
12	59	50	42	53	43	46
13	44	42	42	42	43	47
14	22	21	21	25	18	19
15	8	15	16	14	11	3
16	45	42	37	44	37	34
17	7	14	14	12	13	16
18	23	35	32	34	32	35
19	7	13	13	12	12	12
20	7	12	14	11	5	5
21	20	24	26	22	28	26
22	22	21	23	21	26	28
23	17	17	18	17	12	11
24	16	16	19	16	15	21
25	16	16	20	16	16	15

The cumulative comparison of actual and predicted ROP values versus depth is plotted on Figure 6.101.

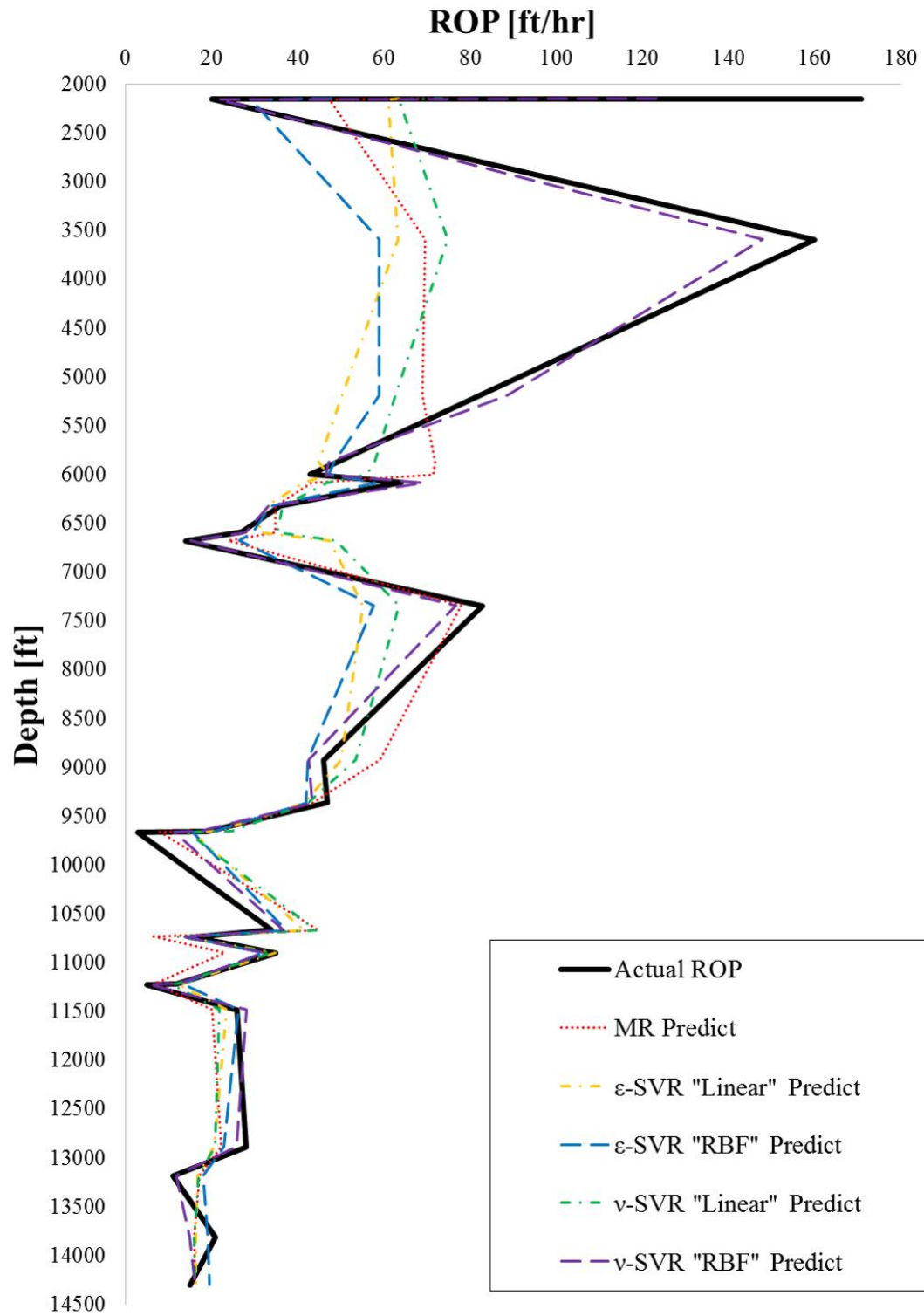


Figure 6.101 : ROP prediction trends in Case 7 for testing all.

The comparison of each prediction with actual ROP is plotted in Figure 6.102.

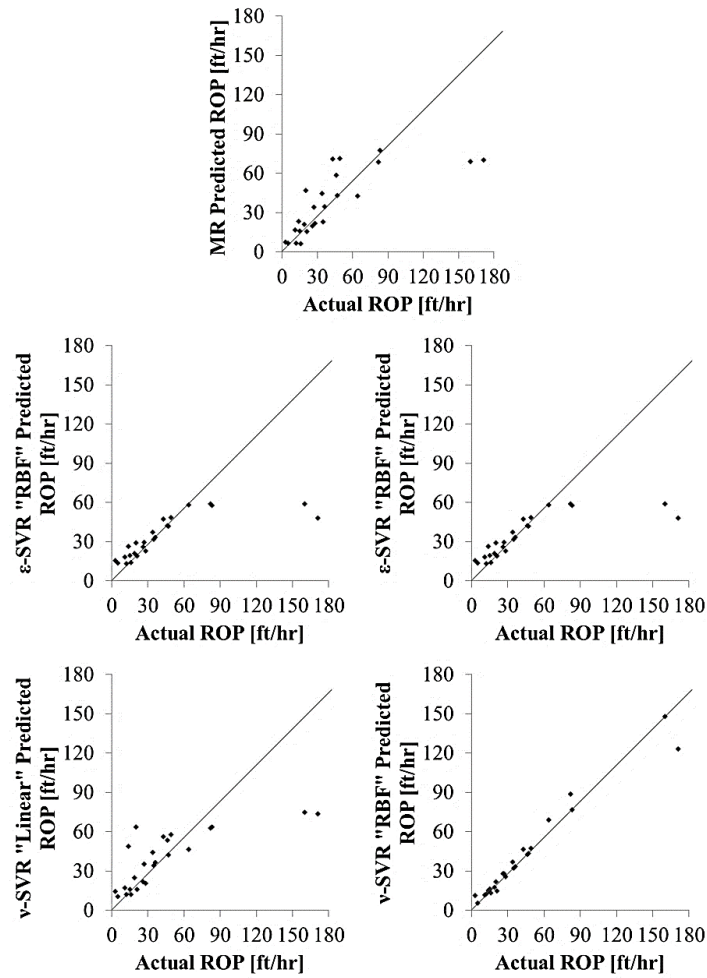


Figure 6.102 : ROP prediction comparison in Case 7 for testing all.

It is seen from Figures 6.101 and 6.102 that, similarly to the previous case, ν -SVR with RBF kernel is the most accurate predictor for the whole data set. Additionally, ν -SVR with RBF kernel has the minimum RMSE, RSS and maximum Pseudo- R^2 values among the other methods.

6.8 Discussions and General Interpretations

It is stated by Flach (2012) that SVM makes better estimation with less training data when compared to the other major machine learning methods. Similarly, it can be seen from the results that SVR can make predictions with good approximation when the number of training data is limited. This situation can be confirmed by the results of first 4 cases. Additionally, in almost all cases, SVR methods generated better predictions when compared to MR.

As mentioned in Chapter 2, some of the machine learning methods, such as SVM, ANN and GA, were used in most of the recent drilling optimization studies. Additionally, ANN and GA were selected particularly for ROP prediction and optimization studies. Moreover, according to Tolun (2008) SVM can make more accurate predictions rather than ANN, especially when there is less number of training data. Similarly, it can be stated that SVM (or SVR) can be preferred as the predictor method when the data is limited.

7. CONCLUSIONS AND RECOMMENDATIONS

As summary, one of the most commonly used machine learning methods, Support Vector Regression (SVR), is applied to the data set taken from the study of Bourgoyne Jr. and Young Jr. (1974) with different implementations and kernel functions. The results of SVR and multiple regression were compared to each other in a statistical point of view to determine the best predictor in each case.

The study is concluded with the following outcomes.

- When the number of training data is insufficient (being less than 30 data set), the accuracy of MR is low. However, if the number of training data increases, MR analysis gives more accurate results, as expected.
- For the scenarios including limited number of training data, SVR methods with linear kernels give more accurate on ROP predictions. Yet, when the number of training data is increased, RBF kernels, especially in ν -SVR, generates better predictions. This outcome can be confirmed by the results of Case 6 and 7.
- The Pseudo- R^2 results are unstable in overall. The number of training data and the selection of test data have a remarkable effect on the prediction accuracy.
- It is considered that the data taken from the study of Bourgoyne Jr. and Young Jr. (1974) is inconsistent in itself, since there are leaps in ROP at some certain points of depth. This situation creates noise in the selected training data set. Thus, a preliminary statistical analysis should be applied to the data set before selecting training data, which is a subject of another possible study.
- There are slight differences between the MR coefficients found in this study and the study of Bourgoyne Jr. and Young Jr. (1974).
- The data collecting technology in 1974 is considered to be not reliable. Hence, the methods should be tested with a field data taken with recent

technology. However, the data number is not enough to generalize on the best predicting method. A data set with more input numbers is needed for further calculations.

- Different implementations of SVR is applied first time in the literature to the data set taken from the study of Bourgoyne Jr. and Young Jr. (1974). The results are data-specific. Thus, the best ROP predicting methods of SVR determined in this study, which are ε -SVR “Linear” for limited data and ν -SVR “RBF” for sufficient data, are not a generalized solution to the ROP optimization problem.
- It is recommended to rely on MR for ROP optimization if there is sufficient amount of data. On the other hand, if the data is limited, such as 50% of the data required for MR, SVR can be used instead of MR to obtain more accurate predictions.
- The investigation on the applicability of SVR for ROP optimization in real-time drilling operations is recommended for further studies.

REFERENCES

- Aadnøy, B. S.** (2010). *Modern Well Design*. Taylor & Francis Group, London, UK.
- Abe, S.** (2010). *Support Vector Machines for Pattern Classification*. Springer, London, UK.
- Adams, N. J.** (1985). *Drilling Engineering: A Complete Well Planning Approach*. PennWell Publishing Company, Tulsa, OK.
- Akgun, F.** (2002). How to estimate the maximum achievable drilling rate without jeopardizing safety, *10th Abu Dhabi International Petroleum Exhibition and Conference*, Abu Dhabi, 13-16 October. doi:10.2118/78567-MS.
- Al-Baiyat, I. and Heinze, L.** (2012). Implementing artificial neural networks and support vector machines in stuck pipe prediction, *SPE Kuwait International Petroleum Conference and Exhibition*, Kuwait City, Kuwait, 10-12 December. doi:10.2118/163370-MS.
- Al-Betairi, E. A., Moussa, M. M., and Al-Otaibi, S.** (1988). Multiple regression approach to optimize drilling operations in the Arabian Gulf area, *SPE Drilling Engineering*, **March 1988**, 83-88. doi:10.2118/13694-PA.
- Alpaydm, E.** (2010). *Introduction to Machine Learning*. The MIT Press, London, UK.
- Aronszajn, N.** (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*. Vol. **68**. no. 3. pp. 337-404.
- Bahari, A. and Baradaran Seyed, A.** (2007). Trust-region approach to find constants of Bourgoyne and Young penetration rate model in Khangiran Iranian gas field, *SPE Latin American and Caribbean Petroleum Engineering Conference*, Buenos Aires, Argentina, 15-18 April. doi:10.2118/107520-MS
- Bahari, A. and Baradaran Seyed, A.** (2009). Drilling cost optimization in a hydrocarbon field by combination of comparative and mathematical methods. *Pet.Sci*, **6**, 451-463. doi:10.1007/s12182-009-0069-x.

- Bahari, M. H., Bahari, A., Nejati Moharrami, F., and Naghibi Sistani, M. B.** (2008). Determining Bourgoyne and Young model coefficients using genetic algorithm to predict drilling rate. *Journal of Applied Sciences*. Vol. 8. no. 17. pp. 3050-3054.
- Barnston, A. G.** (1992). Correspondence among the correlation, RMSE, Heidke forecast verification measures; refinement of the Heidke score. *Weather and Forecasting*. Vol. 7. no 4. pp. 699-709.
- Barragan, R. V., Santos, O. L. A., and Maidla, E.E.** (1997). Optimization of multiple bit runs, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 4-6 March 1997. doi:10.2118/37644-MS.
- Bataee, M. and Mohseni, S.** (2011). Application of artificial intelligent systems in ROP optimization: A case study in Shadegan oil field, *SPE Middle East Unconventional Gas Conference and Exhibition*, Muscat, Oman, 31 January-2 February. doi:10.2118/140029-MS.
- Ben-Hur, A. and Weston, J.** (2009). A user's guide to support vector machines. In O. Carugo and F. Eisenhaber (Ed.), *Data Mining Techniques for the Life Sciences*: Vol. 609. *Methods in molecular biology*, (pp. 223-239). Humana Press. doi:10.1007/978-1-60327-241-4_13.
- Bilgesu, H. I., Tetrick, L. T., Altmis, U., Mohaghegh, S., and Ameri, S.** (1997). A new approach for the prediction of rate of penetration (ROP) values, *SPE Eastern Regional Meeting*, Lexington, KY, 22-24 October. doi:10.2118/39231-MS.
- Bilgin, N., Copur, H., and Balci, C.** (2014). *Mechanical Excavation in Mining and Civil Industries*. Taylor & Francis Group, NW.
- Bingham, M. G.** (1965). *A New Approach to Interpreting: Rock Drillability*. Petroleum Pub. Co.
- Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer, NY.
- Bizanti, M. S. and Blick, E. F.** (1986). Fluid dynamics of wellbore bottomhole cleaning, *Permian Basin Oil & Gas Recovery Conference of the Society of Petroleum Engineers*, Midland, TX, 13-14 March. doi:10.2118/15010-MS.
- Boswell, D.** (2002). Introduction to support vector machines. *Report*, Caltech, CA, USA.

- Bourgoyne Jr., A. T. and Young Jr., F. S.** (1974). A multiple regression approach to optimal drilling and abnormal pressure detection. *Society of Petroleum Engineers Journal*, August, 371-384. doi:10.2118/4238-PA.
- Bourgoyne Jr., A. T., Millheim, K. K., Chenevert, M. E., and Young Jr., F. S.** (1991). *Applied Drilling Engineering*. Society of Petroleum Engineers, Richardson, TX.
- Bousquet, O., Boucheron, S., and Lugosi, G.** (2004). Introduction to statistical learning theory. In O. Bousquet, U. von Luxburg, and G. Rätsch (Ed.), *Advanced Lectures on Machine Learning: Vol. 3176. Lecture notes in computer science* (pp. 169-207). Springer Berlin Heidelberg. doi:10.1007/978-3-540-28650-9_8.
- Burges, C. J. C.** (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 121-167.
- Caicedo, H. U., Calhoun, W. M., and Ewy, R. T.** (2005). Unique ROP predictor using bit-specific coefficient of sliding friction and mechanical efficiency as a function of confined compressive strength impacts drilling performance, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 23-25 February. doi:10.2118/92576-MS.
- Cayeux, E., Dvergsnes E. W., and Iversen, F.** (2009). Real-time optimization of the drilling process – challenges in industrialization, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 17-19 March. doi:10.2118/119650-MS.
- Chang, C. C. and Lin, C. J.** (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, Vol. 3, no. 27. pp. 1-27. doi:10.1145/1961189.1961199.
- Combs, G. D.** (1968). Prediction of pore pressure from penetration rate. *American Institute of Mining, Metallurgical and Petroleum Engineers, Inc.* doi:10.2118/2162-MS.
- Cooper, G. A., Cooper, A. G., and Bihn, G.** (1995). An interactive drilling simulator for teaching and research, *Petroleum Computer Conference*, Houston, TX, 11-14 June. doi:10.2118/30213-MS.
- Cortes, C. and Vapnik, V.** (1995). Support-vector networks. *Machine Learning*, 20, 273-297.

- Cunningham, R. A. and Eenink, J. G.** (1958). Laboratory study of effect of overburden, formation and mud column pressures on drilling rate of permeable formations, *33rd Annual Fall Meeting of Society of Petroleum Engineers*, Houston, TX, 5-8 October.
- Cyril, O., Peter, O., and Ebuefegha, A.** (2014). Using decision algorithm to optimize development of a gas field with limited subsurface data, *Nigeria Annual International Conference and Exhibition*, Lagos, Nigeria, 05-07 August. doi:10.2118/172459-MS.
- Devereux, S.** (2012). *Drilling Technology in Nontechnical Language*. PennWell Corporation, Tulsa, OK.
- Dietterich, T.** (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys*, Vol. 27, no. 3, pp. 326-327.
- Doiron, H. H. and Deane, J. D.** (1982). Effects of hydraulic parameter cleaning variations on rate of penetration of soft formation insert bits, *57th Annual Fall Technical Conference and Exhibition of the Society of Petroleum Engineers of AIME*, New Orleans, LA, 26-29 September. doi:10.2118/11058-MS.
- Dubinsky, V. S. and Baecker, D. R.** (1998). An interactive drilling dynamics simulator for drilling optimization and training, *SPE Annual Technical Conference and Exhibition*, New Orleans, LA, 27-30 September. doi:10.2118/49205-MS.
- Dufour, J. M.** (2011). Coefficients of determination. *Report*, Department of Economics, McGill University, Canada.
- Dupriest, F. E. and Koederitz, W. L.** (2005). Maximizing drill rates with real-time surveillance of mechanical specific energy, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 23-25 February. doi:10.2118/92194-MS.
- Eckel, J. R.** (1967). Microbit studies of the effect of fluid properties and hydraulics on drilling rate. *Journal of Petroleum Technology*, April, 542-546. doi:10.2118/1520-PA.
- Edwards, J. H.** (1964). Engineering design of drilling operations. *Spring Meeting of the Southern District, API Division of Production*, March.

- Eren, T. and Ozbayoglu, M. E.** (2010). Real time optimization of drilling parameters during drilling operations, *SPE Oil and Gas India Conference and Exhibition*, Mumbai, India, 20-22 January. doi:10.2118/129126-MS.
- Eren, T.** (2010). Real-time-optimization of drilling parameters during drilling operations, *PhD Thesis*, METU, Ankara.
- Esmacili, A., Elahifar, B., Fruhwirth, R. K., and Thonhauser, G.** (2012). ROP modeling using neural network and drill string vibration data, *SPE Kuwait International Petroleum Conference and Exhibition*, Kuwait City, Kuwait, 10-12 December. doi:10.2118/163330-MS.
- Fear, M. J.** (1996). How to improve rate of penetration in field operations, *IADC/SPE Drilling Conference*, New Orleans, LA, 12-15 March. doi:10.2118/35107-MS.
- Flach, P.** (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press, Cambridge, UK.
- Fletcher, R.** (2000). *Practical Methods of Optimization*. John Wiley & Sons.
- Fonticella, R.** (n.d.). The usefulness of the R^2 statistic. CAS Publications.
- Freund, R. J., Wilson, W. J., and Sa, P.** (2006). *Regression Analysis: Statistical Modeling of a Response Variable*. Elsevier, Burlington, MA.
- Galle, E. M. and Woods, H. B.** (1963). Best constant weight and rotary speed for rotary rock bits, *Spring Meeting of the Pacific Coast District, API Division of Production*, May.
- Garnier, A. J. and van Lingen, N. H.** (1959). Phenomena affecting drilling rates at depth. *Petroleum Transactions, AIME*, Vol. **216**, pp. 232-239.
- Gelfgat, Y. A., Gelfgat, M. Y., and Lopatin, Y. S.** (2003). *Advanced Drilling Solutions: Lessons from the FSU*. PennWell Corporation, Tulsa, OK.
- Ghahramani, Z.** (2004). Unsupervised learning. In O. Bousquet, U. von Luxburg, and G. Rätsch (Ed.), *Advanced Lectures on Machine Learning: Vol. 3176. Lecture notes in computer science* (pp. 72-112). Springer Berlin Heidelberg. doi:10.1007/978-3-540-28650-9_5.
- Gidh, Y., Purwanto, A., and Ibrahim, H.** (2012). Artificial neural network drilling parameter optimization system improves ROP by predicting/managing bit wear, *SPE Intelligent Energy International*, Utrecht, The Netherlands, 27-29 March.

- Girosi, F.** (1997). An equivalence between sparse approximation and support vector machines, *Massachusetts Institute of Technology Artificial Intelligence Laboratory & Center for Biological and Computational Learning Department of Brain and Cognitive Sciences Technical Report, AI Memo No. 1606, CBCL Paper No. 147*, Massachusetts Institute of Technology, Cambridge, MA.
- Graham, J. W. and Muench, N. L.** (1959). Analytical determination of optimum bit weight and rotary speed combinations, *Fall Meeting of the Society of Petroleum Engineers of AIME*, Dallas, TX, 4-7 October.
- Gunn, S. R.** (1998). Support vector machines for classification and regression, *School of Electronics and Computer Science Technical Report*, University of Southampton, UK.
- Hareland G., Wu, G., Rashidi, B., and James, J. A.** (2010). A new drilling rate model for tricone bits and its application to predict rock compressive strength. *44th US Rock Mechanics Symposium*, Salt Lake City, UT, 27-30 June.
- Heckman, N. E.** (1997). The theory and application of penalized least squares methods or reproducing kernel Hilbert spaces made easy. Preprint.
- Hou, B., Chen, M., and Yuan, J.** (2014). Optimization and application of bit selection technology for improving the penetration rate. *Research Journal of Applied Sciences, Engineering and Technology*, Vol. 8, no: 2, pp. 179-187.
- Huber, P. J.** (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, Vol. 35, no. 1, pp.73-101.
- Ipek, G., Smith, J. R., and Bassiouni, Z.** (2006). Diagnosis of ineffective drilling using cation exchange capacity of shaly formations. *Journal of Canadian Petroleum Technology*, Vol. 45, no. 8, pp. 26-30.
- Iqbal, F.** (2008). Drilling optimization technique – using real time parameters, *SPE Russian Oil & Gas Technical Conference and Exhibition*, Moscow, Russia, 28-30 October. doi:10.2118/114543-MS.
- Iqbal, K.** (2013). *Fundamental Engineering Optimization Methods*. Bookboon.
- Irawan, S., Abd Rahman, A. M., and Tunio, S. Q.** (2012). Optimization of weight on bit during drilling operation based on rate of penetration model. *Research Journal of Applied Sciences, Engineering and Technology*, Vol. 4, no. 12, pp. 1690-1695.

- Irawan, S. and Anwar, I.** (2012). Optimization of weight on bit during drilling operation based on rate of penetration model. *Jurnal Aptek*, Vol. 4, no. 1, pp. 55-64.
- Iversen, F. P., Cayeux, E., Dvergsnes, E. W., Gravdal, J. E., Vefring, E. H., et al.** (2006). Monitoring and control of drilling utilizing continuously updated process models, *IADC/SPE Drilling Conference*, Miami, FL, 21-23 February. doi:10.2118/99207-MS.
- Iversen, F. P., Cayeux, E., Dvergsnes, E. W., Ervik, R., Byrkjeland, M., et al.** (2008). Offshore field test of a new integrated system for real-time optimization of the drilling process, *IADC/SPE Drilling Conference*, Miami, FL, 4-6 March. doi:10.2118/112744-MS.
- Jacinto, C. M. C., Freitas Filho, P. J., Nassar, S. M., Roisenberg, M., et al.** (2013). Optimization models and prediction of drilling rate (ROP) for the Brazilian pre-salt layer. *Chemical Engineering Transactions*, Vol. 33, pp. 823-828. doi:10.3303/CET1333138.
- Jamshidi, E. and Mostafavi, H.** (2013). Soft computation application to optimize drilling bit selection utilizing virtual intelligence and genetic algorithms, *International Technology Conference*, Beijing, China, 26-28 March. doi:10.2523/16446-MS.
- Jeng, J. T. and Chuang, C. C.** (2002). A novel approach for the hyperparameters of support vector regression. *Proceedings of IJCNN '02*, Vol. 1, pp. 642-647. doi:10.1109/IJCNN.2002.1005547.
- John, Z., Ahsan, A., and Reid, I.** (2002). Optimized decision making through real time access to drilling and geological data from remote wellsites, *SPE Asia Pacific Oil and Gas Conference and Exhibition*, Melbourne, Australia, 8-10 October. doi:10.2118/77855-MS.
- Judzis, A., Bland, R. G., Curry, D. A., Black, A. D., Robertson, H. A., et al.** (2007). Optimization of deep drilling performance: Benchmark testing drives ROP improvements for bits and drilling fluids, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 20-22 February. doi:10.2118/105885-MS.
- Kecman, V.** (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. The MIT Press, London, UK.

- Kulkarni, S. R. and Harman, G.** (2011). Statistical learning theory: A tutorial. *WIREs Comp Stat*, **3**: 543–556. doi: 10.1002/wics.179.
- Lagenkamp, R. D.** (1994). *Handbook of Oil Industry Terms and Phrases*. PennWell Publishing Company, Tulsa, OK.
- Lubinski, A.** (1988). *Developments in Petroleum Engineering: Volume 2*. Gulf Publishing Company, Houston, TX.
- Lummus, J. L.** (1970). Drilling optimization. *Journal of Petroleum Technology*, November, pp. 1379-1388.
- Lummus, J. L.** (1971). Acquisition and analysis of data for optimized drilling. *Journal of Petroleum Technology*, November, pp. 1285-1293.
- Lyons, W.** (2010). *Working Guide to Drilling Equipment and Operations*. Elsevier, Burlington, MA.
- Maidla, E. E. and Ohara, S.** (1991). Field verification of drilling models and computerized selection of drill bit, WOB, and drillstring rotation. *SPE Drilling Engineering*, pp. 189-195.
- Malhotra, Y.** (2001). Expert systems for knowledge management: Crossing the chasm between information processing and sense making, *Expert Systems with Application*, Vol. **20**, pp. 7-16.
- Maurer, W. C.** (1962). The “perfect-cleaning” theory of rotary drilling. *Journal of Petroleum Technology*, November, pp. 1270-1274.
- Mercer, M. A.** (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, Vol. **209**, pp. 415-446.
- Millheim, K. K. and Gaebler, T.** (1999). Virtual experience simulation for drilling – the concept, *SPE/IADC Drilling Conference*, Amsterdam, Holland, 9-11 March. doi:10.2118/52803-MS.
- Milner, J., Bergjord, O. G., Høyland, K. and Rugland, B.** (2006). Use of real-time data at the Statfjord field anno 2005, *SPE Intelligent Energy Conference and Exhibition*, Amsterdam, The Netherlands, 11-13 April. doi:10.2118/99257-MS.
- Mitchell, B.** (1995). *Advanced Oilwell Drilling Engineering Handbook & Computer Programs*. Mitchell Engineering, Lakewood, CO.

- Mittlböck, M. and Heinzl, H.** (2004). Pseudo r-squared measures for generalized linear models. *1st European Workshop on the Assessment of Diagnostic Performance Proceedings*, 71-80.
- Miyora, T., Jónsson, M. Þ., Þórhallsson, S.** (2015). Modelling and optimization of geothermal drilling parameters – a case study of well MW-17 in Menengai Kenya, *40th Workshop on Geothermal Reservoir Engineering*, Stanford, CA, 26-28 January.
- Mochizuki, S., Saputelli, L. A., Kabir, C. D., Cramer, R., Lochmann, M. J., et al.** (2004). Real time optimization: Classification and assessment, *SPE Annual Technical Conference and Exhibition*, Houston, TX, 26-29 September. doi:10.2118/90213-PA.
- Monazami, M., Hashemi, A., and Shahbazian, M.** (2012). Drilling rate of penetration prediction using artificial neural network: A case study of one of Iranian southern oil fields. *Oil and Gas Business*, no. 6, pp. 21-31.
- Monden, T. and Chia, C. R.** (2007). Operation support centers – real time drilling optimization and risk mitigation, *SPE Saudi Arabia Technical Symposium*, Dhahran, Saudi Arabia, 7-8 May. doi:10.2118/110950-MS.
- Montgomery, D. C. and Runger, G. C.** (2003). *Applied Statistics and Probability for Engineers*. John Wiley & Sons, New York, NY.
- Montgomery, D. C., Peck, E. A., and Vining, G.** (2006). *Introduction to Linear Regression Analysis*. John Wiley & Sons, New York, NY.
- Moore, P. L.** (1986). *Drilling Practices Manual*. PennWell Publishing Company, Tulsa, OK.
- Moradi, H., Bahari, M. H., Sistani, M. B. N., and Bahari, A.** (2010). Drilling rate prediction using an innovative soft computing approach. *Scientific Research and Essays*, Vol. 5, no. 13, pp. 1583-1588.
- Mulmuley, K.** (1994). *Computational Geometry: An Introduction Through Randomized Algorithms*. Prentice Hall, Englewood Cliffs, NJ.
- Murray, A. S. and Cunningham, R. A.** (1955). Effect of mud column pressure on drilling rates, *Petroleum Branch Fall Meeting*, New Orleans, LA, 2-5 October.
- Negnevitsky, M.** (2002). *Artificial Intelligence: A Guide to Intelligent Systems*. Pearson Education Limited, Essex, UK.

- Nocedal, J. and Wright, S. J.** (1999). *Numerical Optimization*. Springer, New York, NY.
- Okay, A. I.** (2008). Geology of Turkey: A synopsis. *Anschnitt*, Vol. **21**, pp. 19-42.
- Osgouei, R. E.** (2007). Rate of penetration estimation model for directional and horizontal wells, *MSc Thesis*, METU, Ankara.
- Pessier, R. C. and Fear, M. J.** (1992). Quantifying common drilling problems with mechanical specific energy and a bit-specific coefficient of sliding friction, *67th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers*, Washington, DC, 4-7 October. doi:10.2118/24584-MS.
- Rahimzadeh, H., Mostofi, M., and Hashemi, A.** (2011). A new method for determining Bourgoyne and Young penetration rate model constants. *Petroleum Science and Technology*, Vol. **29**, no. 9, pp. 886-697.
- Rampersad, P. R., Hareland, G., and Boonyapaluk, P.** (1994). Drilling optimization using drilling data and available technology, *III Latin American/Caribbean Petroleum Engineering Conference*, Buenos Aires, Argentina, 27-29 April. doi:10.2118/27034-MS.
- Rao, S. S.** (2009). *Engineering Optimization: Theory and Practice*. John Wiley & Sons, Hoboken, NJ.
- Reed, R. L.** (1972). A Monte Carlo approach to optimal drilling, *SPE 46th Annual Fall Meeting*, New Orleans, LA, 3-6 October.
- Remmert, S. M., Witt, J. W., and Dupriest, F. E.** (2007). Implementation of ROP management process in Qatar north field, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 20-22 February. doi:10.2118/105521-MS.
- Reza, M. R. and Alcocer, C. F.** (1986). A unique computer simulation model well drilling: Part I – the Reza drilling model, *56th California Regional Meeting of the Society of Petroleum Engineers*, Oakland, CA, 2-4 April. doi:10.2118/15108-MS.
- Rommetveit, R., Bjørkevoll, K. S., Halsey, G. W., Larsen, H. F., Merlo, A., et al.** (2004). Drilltronics: An integrated system for real-time optimization of the drilling process, *IADC/SPE Drilling Conference*, Dallas, TX, 2-4 March. doi:10.2118/87124-MS.

- Samuel, G. R. and Miska, S.** (1998). Optimization of drilling parameters with the performance of multilobe positive displacement motor (PDM), *IADC/SPE Asia Pacific Drilling Conference*, Jakarta, Indonesia, 7-9 September. doi:10.2118/47791-MS.
- Schölkopf, B., Bartlett, P., Smola, A., and Williamson, R.** (1998). Support vector regression with automatic accuracy control. In L. Niklasson, M. Bodén, and T. Ziemke (Ed.), *ICANN 98, Perspectives in neural computing* (pp. 111-116). Springer London. doi:10.1007/978-1-4471-1599-1_12.
- Serpen, U., Alpkaya, E., Yamanlar, Ş., and Özkara, H.** (1997). Sondaj optimizasyonu için bilgisayar programları geliştirilmesi, TP, Ankara.
- Shi, Z.** (2011). *Advanced Artificial Intelligence*. World Scientific, London, UK.
- Shishavan, R. A., Hubbell, C., Perez, H. D., Hedengren, J. D., and Pixton, D. S.** (2014). Combined rate of penetration and pressure regulation for drilling optimization using high speed telemetry, *SPE Deepwater Drilling and Completions Conference*, Galveston, TX, 10-11 September. doi:10.2118/170275-MS.
- Simmons, E. L.** (1986). A technique for accurate bit programming and drilling performance optimization, *IADC/SPE Drilling Conference*, Dallas, TX, 10-12 February. doi:10.2118/14784-MS.
- Sivanandam, S. N. and Deepa, S. N.** (2008). *Introduction to Genetic Algorithms*. Springer, New York, NY.
- Sivanandam, S. N., Sumathi, S., and Deepa, S. N.** (2007). *Introduction to Fuzzy Logic using MATLAB*. Springer, New York, NY.
- Smith Bits.** (2008). *Roller Cone Dull Grading Manual*. Smith International, Houston, TX.
- Smola, A. J. and Schölkopf, B.** (2003). A tutorial on support vector regression. *Statistics and Computing*, Vol. 14, no. 3, pp. 199-222. doi:10.1023/B:STCO.0000035301.49549.88.
- Smola, A. J., Murata, N., Schölkopf, B., and Müller, K. R.** (1998). Asymptotically optimal choice of ϵ -loss for support vector machines. In L. Niklasson, M. Bodén, and T. Ziemke (Ed.), *ICANN 98, Perspectives in neural computing* (pp. 105-110). Springer London. doi:10.1007/978-1-4471-1599-1_11.

- Smola, A.** (1996). Regression estimation with support vector learning machines, *Mphys Thesis*, Technische Universität München, Germany.
- Speer, J. W.** (1958). A methods for determining optimum drilling techniques, *Spring Meeting of the Southern District, API Division of Production*, Houston, TX, February.
- Strathman, M., Elley, D., and Meirerhoefer, N.** (2007). Time-based real-time-drilling-operations excellence delivered, *SPE Digital Energy Conference and Exhibition*, Houston, TX, 11-12 April. doi:10.2118/107303-MS.
- Tansev, E.** (1975). A heuristic approach to drilling optimization, *50th Annual Fall Meeting of the Society of Petroleum Engineers of AIME*, Dallas, TX, 28 October. doi:10.2118/5546-MS.
- Teale, R.** (1965). The concept of specific energy in rock drilling. *Int. J. Rock Mech. Mining ScL*. Vol. 2, pp. 57-73.
- Tibbitts, G. A., Sandstrom, J. L., Black, A. D., and Green, S. J.** (1981). Effects of bit hydraulics on full-scale laboratory drilled shale. *Journal of Petroleum Technology*, July, pp. 1180-1188.
- Tollefsen, E., Goobie, R. B., Noeth, S., Sayers, C., den Boer, L., Hooyman, P., et al.** (2006). Optimize drilling and reduce casing strings using remote real-time well hydraulic, *First International Oil Conference and Exhibition in Mexico*, Cancun, Mexico, 31 August-2 September. doi:10.2118/103936-PA.
- Tolun, S.** (2008). Destek vektör makineleri: Banka başarısızlığının tahmini üzerine bir uygulama, *PhD Thesis*, Istanbul University, Istanbul.
- Ursem, L. J., Williams, J. H., Pellerin, N. M., and Kaminski, D. H.** (2003). Real time operations centers; the people aspects of drilling decision making, *SPE/IADC Drilling Conference*, Amsterdam, The Netherlands, 19-21 February. doi:10.2118/79893-MS.
- Vapnik, V.** (1991). Principles of risk minimization for learning theory. In J. E. Moody, S. J. Hanson, and R. P. Lipmann (Ed.), *Advances in Neural Information Processing Systems: Vol. 4. Neural Information Processing Systems* (pp. 831-838). Neural Information Processing Systems Foundation, Inc.
- Vapnik, V. N.** (1998). *Statistical Learning Theory*. John Wiley & Sons, Danvers, MA.

- Vapnik, V. N.** (1999). *The Nature of Statistical Learning Theory*. Springer, New York, NY.
- Vidrine, D. J. and Benit, E. J.** (1968). Field verification of the effect of differential pressure on drilling rate. *Journal of Petroleum Technology*, July, pp. 676-682.
- Wahba, G.** (1990). *Spline Models for Observational Data*. SIAM, Philadelphia, PA.
- Wardlaw, H. W. R.** (1969). Drilling performance optimization and identification of overpressure formations. SPE 2388. Preprint.
- Warren, T. M.** (1987). Penetration-rate performance of roller-cone bits. *SPE Drilling Engineering*, March, pp. 9-18.
- Wilson, D. C. and Bentsen, R. G.** (1972). Optimization techniques for minimizing drilling costs, *47th Annual Fall Meeting of the Society of Petroleum Engineers of AIME*, San Antonio, TX, 8-11 October. doi:10.2118/3983-MS.
- Witten, I. H. and Frank, E.** (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, San Francisco, CA.
- Wojtanowicz, A. K. and Kuru, E.** (1993). Minimum-cost well drilling strategy using dynamic programming. *Journal of Energy Resources Technology*, Vol. *115*, pp. 239-246.
- Wright, J., Chukwu, G. A., Khataniar, S., and Patil, S.** (2003). An economic appraisal of hole cleaning using hydraulic horsepower and jet impact force, *SPE Western Regional/AAPG Pacific Section Joint Meeting*, Long Beach, CA, 19-24 May. doi:10.2118/83496-MS.
- Wu, A., Hareland, G., and Rashidi, B.** (2010). The effect of different rock types and roller cone insert types and wear on ROP (rate of penetration). *44th US Rock Mechanics Symposium*, Salt Lake City, UT, 27-30 June.
- Yan, X. and Su, X. G.** (2009). *Linear Regression Analysis: Theory and Computing*. World Scientific, Singapore.
- Yen, T. F. and Chilingarian, G. V.** (1976). *Oil Shales*. Elsevier, New York, NY.
- Young Jr., F. S.** (1969). Computerized drilling control. *Journal of Petroleum Technology*, April, pp. 483-496.
- Zhang, T.** (2001). An introduction to support vector machines and other kernel-based learning methods: A review. *AI Magazine*, Vol. *22*, no. 2, pp. 103-104.

- Meyer, D.** (2014). Support vector machines: The interface to libsvm in package e1071. Date retrieved: 07.04.2015, address: <http://http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>
- Venables, W. N. and Smith, D. M.** (2014). *An Introduction to R* (User manual). Retrieved from <http://http://cran.r-project.org/doc/manuals/R-intro.pdf>
- Williams, R.** (2015). *Scalar measures of fit: Pseudo R^2 and information measures (AIC & BIC)* (Lecture notes). Retrieved from <http://www3.nd.edu/~rwilliam/stats3/L05.pdf>

APPENDICES

APPENDIX A: Data Sets and Multiple Regression Solution

APPENDIX B: Program Code

APPENDIX A

Table A.1 : The main data set (Bourgoyne Jr. and Young Jr., 1974, p. 375).

Data Entry	Depth [ft]	ROP [ft/hr]	Bit Weight [1.000 lb/in]	Rotary Speed [rpm]	Tooth Wear	Reynolds Number Function	ECD [lb/gal]	Pore Gradient [lb/gal]
1	9.515	23,0	2,58	113	0,77	0,964	9,5	9,0
2	9.830	22,0	1,15	126	0,38	0,964	9,5	9,0
3	10.130	14,0	0,81	129	0,74	0,827	9,6	9,0
4	10.250	10,0	0,95	87	0,15	0,976	9,7	9,0
5	10.390	16,0	1,02	78	0,24	0,984	9,7	9,0
6	10.500	19,0	1,69	81	0,61	0,984	9,7	9,1
7	10.575	13,0	1,56	81	0,73	0,984	9,7	9,2
8	10.840	16,6	1,63	67	0,38	0,932	9,8	9,3
9	10.960	15,9	1,83	65	0,57	0,878	9,8	9,4
10	11.060	15,7	2,03	69	0,72	0,878	9,8	9,5
11	11.475	14,0	1,69	77	0,20	0,887	10,3	9,5
12	11.775	13,5	2,31	58	0,12	0,852	11,8	10,1
13	11.940	6,2	2,26	67	0,2	0,976	15,3	12,4
14	12.070	9,6	2,07	84	0,08	0,993	15,7	13,0
15	12.315	15,5	3,11	69	0,40	1,185	16,3	14,4
16	12.900	31,4	2,82	85	0,42	1,150	16,7	15,9
17	12.975	42,7	3,48	77	0,17	1,221	16,7	16,1
18	13.055	38,6	3,29	75	0,29	1,161	16,8	16,2
19	13.250	43,4	2,82	76	0,43	1,161	16,8	16,2
20	13.795	12,5	1,60	81	0,56	0,272	16,8	16,2
21	14.010	21,1	1,04	75	0,46	0,201	16,8	16,2
22	14.455	19,0	1,76	64	0,16	0,748	16,9	16,2
23	14.695	18,7	2,00	76	0,27	0,819	17,1	16,2
24	14.905	20,2	2,35	75	0,33	0,419	17,2	16,4
25	15.350	27,1	2,12	85	0,31	1,29	17,0	16,5
26	15.740	14,8	2,35	78	0,81	0,802	17,3	16,5
27	16.155	12,6	2,47	80	0,12	0,670	17,9	16,5
28	16.325	14,9	3,76	81	0,50	0,532	17,5	16,6
29	17.060	13,8	3,76	65	0,91	0,748	17,6	16,6
30	20.265	9,0	3,41	60	0,01	0,512	17,7	16,6

Table A.2 : The recent data set (Irawan and Anwar, 2012, p. 62).

Data Entry	Depth [ft]	ROP [ft/hr]	Bit Weight [1.000 lb/in]	Rotary Speed [rpm]	Tooth Wear	Reynolds Number Function	ECD [lb/gal]	Pore Gradient [lb/gal]
1	2150	171	0,82	120	0,500	0,882	8,93	8,365
2	2155	20	0,57	110	0,125	0,819	9,06	8,365
3	3591	160	0,82	120	0,500	1,290	9,11	8,365
4	5190	82	1,63	120	0,750	1,290	9,11	8,365
5	5872	49	2,45	120	0,875	1,290	9,11	8,365
6	6000	43	2,45	120	0,250	1,290	9,11	8,365
7	6080	64	1,63	120	0,625	1,062	9,49	8,365
8	6322	36	2,45	120	0,875	0,772	9,67	8,365
9	6592	27	2,85	120	1,000	0,772	9,67	8,365
10	6679	14	0,41	120	0,625	1,338	9,69	8,365
11	7341	83	1,63	180	0,375	1,145	9,69	8,365
12	8921	46	1,63	180	0,000	1,216	9,68	8,365
13	9363	47	1,63	180	0,000	0,868	9,88	8,571
14	9652	19	2,85	100	1,000	1,192	9,96	8,960
15	9660	3	2,45	65	0,125	1,192	9,96	8,960
16	10662	34	1,22	180	0,000	1,097	9,96	8,910
17	10735	16	2,86	65	0,125	1,192	9,96	8,900
18	10900	35	0,82	150	0,000	1,034	9,96	8,890
19	11214	12	3,27	70	0,250	1,114	9,96	8,880
20	11224	5	2,94	100	0,375	0,903	11,10	9,390
21	11481	26	1,76	170	0,000	0,975	11,02	9,370
22	12885	28	1,76	160	0,000	0,975	11,02	9,860
23	13180	11	1,76	130	0,000	0,825	10,96	10,120
24	13810	21	1,76	150	0,000	0,632	10,97	10,040
25	14300	15	1,76	160	0,000	0,632	10,95	9,980

The solution of multiple regression analysis for the data taken from the study of Bourgoyne Jr. and Young Jr. (1974) is given below explicitly.

The ROP model is defined as,

$$R = \exp \left(a_1 + \sum_{j=2}^8 a_j x_j \right) \quad (\text{A.1.1})$$

$$x_2 = 10000 - D \quad (\text{A.1.2})$$

$$x_3 = D^{0.69} (g_p - 9) \quad (\text{A.1.3})$$

$$x_4 = D (g_p - \rho_c) \quad (\text{A.1.4})$$

$$x_5 = \ln \left(\frac{\frac{W_b}{d_b} - \left(\frac{W_b}{d_b} \right)_t}{4 - \left(\frac{W_b}{d_b} \right)_t} \right) \quad (\text{A.1.5})$$

$$x_6 = \ln\left(\frac{N}{100}\right) \quad (\text{A.1.6})$$

$$x_7 = -h \quad (\text{A.1.7})$$

$$x_8 = \ln\left(\frac{F_j}{1000}\right) \quad (\text{A.1.8})$$

Table A.3 : Values of parameter functions.

y	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
3,1	485	0	-4757,5	-0,439	0,122	-0,77	-0,037
3,1	170	0	-4915	-1,247	0,231	-0,38	-0,037
2,6	-130	0	-6078	-1,597	0,255	-0,74	-0,190
2,3	-250	0	-7175	-1,438	-0,139	-0,15	-0,024
2,8	-390	0	-7273	-1,366	-0,248	-0,24	-0,016
2,9	-500	59,514	-6300	-0,862	-0,211	-0,61	-0,016
2,6	-575	119,614	-5287,5	-0,942	-0,211	-0,73	-0,016
2,8	-840	182,512	-5420	-0,898	-0,400	-0,38	-0,070
2,8	-960	245,205	-4384	-0,782	-0,431	-0,57	-0,130
2,8	-1060	308,433	-3318	-0,678	-0,371	-0,72	-0,130
2,6	-1475	316,373	-9180	-0,862	-0,261	-0,20	-0,120
2,6	-1775	708,526	-20017,5	-0,549	-0,545	-0,12	-0,160
1,8	-1940	2211,118	-34626	-0,571	-0,400	-0,20	-0,024
2,3	-2070	2620,825	-32589	-0,659	-0,174	-0,08	-0,007
2,7	-2315	3587,513	-23398,5	-0,252	-0,371	-0,40	0,170
3,4	-2900	4733,212	-10320	-0,350	-0,163	-0,42	0,140
3,8	-2975	4889,928	-7785	-0,139	-0,261	-0,17	0,200
3,7	-3055	4979,876	-7833	-0,195	-0,288	-0,29	0,149
3,8	-3250	5031,083	-7950	-0,350	-0,274	-0,43	0,149
2,5	-3795	5172,976	-8277	-0,916	-0,211	-0,56	-1,302
3,0	-4010	5228,472	-8406	-1,347	-0,288	-0,46	-1,604
2,9	-4455	5342,506	-10118,5	-0,821	-0,446	-0,16	-0,290
2,9	-4695	5403,554	-13225,5	-0,693	-0,274	-0,27	-0,200
3,0	-4905	5608,294	-11924	-0,532	-0,288	-0,33	-0,870
3,3	-5350	5800,642	-7675	-0,635	-0,163	-0,31	0,255
2,7	-5740	5901,936	-12592	-0,532	-0,248	-0,81	-0,221
2,5	-6155	6008,874	-22617	-0,482	-0,223	-0,12	-0,400
2,7	-6325	6133,132	-14692,5	-0,062	-0,211	-0,50	-0,631
2,6	-7060	6322,359	-17060	-0,062	-0,431	-0,91	-0,290
2,2	-10265	7119,816	-22291,5	-0,160	-0,511	-0,01	-0,669

The equation (A.1.1) is transformed into,

$$y = a_1 + a_2x_2 + a_3x_3 + \dots + a_8x_8 \quad (\text{A.1.9})$$

where,

$$y = \ln(R) \quad (\text{A.1.10})$$

Then, the eight least squares equations are solved simultaneously for the correlation coefficients, a_1 through a_8 , for $\theta=30$.

$$\sum y = a_1 \theta + a_2 \sum x_2 + a_3 \sum x_3 + \cdots + a_8 \sum x_8 \quad (\text{A.1.11})$$

$$\sum yx_2 = a_1 \sum x_2 + a_2 \sum x_2x_2 + a_3 \sum x_3x_2 + \cdots + a_8 \sum x_8x_2 \quad (\text{A.1.12})$$

$$\sum yx_3 = a_1 \sum x_3 + a_2 \sum x_2x_3 + a_3 \sum x_3x_3 + \cdots + a_8 \sum x_8x_3 \quad (\text{A.1.13})$$

$$\sum yx_4 = a_1 \sum x_4 + a_2 \sum x_2x_4 + a_3 \sum x_3x_4 + \cdots + a_8 \sum x_8x_4 \quad (\text{A.1.14})$$

$$\sum yx_5 = a_1 \sum x_5 + a_2 \sum x_2x_5 + a_3 \sum x_3x_5 + \cdots + a_8 \sum x_8x_5 \quad (\text{A.1.15})$$

$$\sum yx_6 = a_1 \sum x_6 + a_2 \sum x_2x_6 + a_3 \sum x_3x_6 + \cdots + a_8 \sum x_8x_6 \quad (\text{A.1.16})$$

$$\sum yx_7 = a_1 \sum x_7 + a_2 \sum x_2x_7 + a_3 \sum x_3x_7 + \cdots + a_8 \sum x_8x_7 \quad (\text{A.1.17})$$

$$\sum yx_8 = a_1 \sum x_8 + a_2 \sum x_2x_8 + a_3 \sum x_3x_8 + \cdots + a_8 \sum x_8x_8 \quad (\text{A.1.18})$$

The solution is found via Cramer's rule.

$$\Delta = \begin{vmatrix} \theta & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 & \sum x_6 & \sum x_7 & \sum x_8 \\ \sum x_2 & \sum x_2x_2 & \sum x_3x_2 & \sum x_4x_2 & \sum x_5x_2 & \sum x_6x_2 & \sum x_7x_2 & \sum x_8x_2 \\ \sum x_3 & \sum x_2x_3 & \sum x_3x_3 & \sum x_4x_3 & \sum x_5x_3 & \sum x_6x_3 & \sum x_7x_3 & \sum x_8x_3 \\ \sum x_4 & \sum x_2x_4 & \sum x_3x_4 & \sum x_4x_4 & \sum x_5x_4 & \sum x_6x_4 & \sum x_7x_4 & \sum x_8x_4 \\ \sum x_5 & \sum x_2x_5 & \sum x_3x_5 & \sum x_4x_5 & \sum x_5x_5 & \sum x_6x_5 & \sum x_7x_5 & \sum x_8x_5 \\ \sum x_6 & \sum x_2x_6 & \sum x_3x_6 & \sum x_4x_6 & \sum x_5x_6 & \sum x_6x_6 & \sum x_7x_6 & \sum x_8x_6 \\ \sum x_7 & \sum x_2x_7 & \sum x_3x_7 & \sum x_4x_7 & \sum x_5x_7 & \sum x_6x_7 & \sum x_7x_7 & \sum x_8x_7 \\ \sum x_8 & \sum x_2x_8 & \sum x_3x_8 & \sum x_4x_8 & \sum x_5x_8 & \sum x_6x_8 & \sum x_7x_8 & \sum x_8x_8 \end{vmatrix} \quad (\text{A.1.19})$$

$$a_6 = \frac{\begin{vmatrix} \theta & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 & \sum x_1 & \sum x_7 & \sum x_8 \\ \sum x_2 & \sum x_2 x_2 & \sum x_3 x_2 & \sum x_4 x_2 & \sum x_5 x_2 & \sum x_1 x_2 & \sum x_7 x_2 & \sum x_8 x_2 \\ \sum x_3 & \sum x_2 x_3 & \sum x_3 x_3 & \sum x_4 x_3 & \sum x_5 x_3 & \sum x_1 x_3 & \sum x_7 x_3 & \sum x_8 x_3 \\ \sum x_4 & \sum x_2 x_4 & \sum x_3 x_4 & \sum x_4 x_4 & \sum x_5 x_4 & \sum x_1 x_4 & \sum x_7 x_4 & \sum x_8 x_4 \\ \sum x_5 & \sum x_2 x_5 & \sum x_3 x_5 & \sum x_4 x_5 & \sum x_5 x_5 & \sum x_1 x_5 & \sum x_7 x_5 & \sum x_8 x_5 \\ \sum x_6 & \sum x_2 x_6 & \sum x_3 x_6 & \sum x_4 x_6 & \sum x_5 x_6 & \sum x_1 x_6 & \sum x_7 x_6 & \sum x_8 x_6 \\ \sum x_7 & \sum x_2 x_7 & \sum x_3 x_7 & \sum x_4 x_7 & \sum x_5 x_7 & \sum x_1 x_7 & \sum x_7 x_7 & \sum x_8 x_7 \\ \sum x_8 & \sum x_2 x_8 & \sum x_3 x_8 & \sum x_4 x_8 & \sum x_5 x_8 & \sum x_1 x_8 & \sum x_7 x_8 & \sum x_8 x_8 \end{vmatrix}}{\Delta} \quad (\text{A.1.25})$$

$$a_7 = \frac{\begin{vmatrix} \theta & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 & \sum x_6 & \sum x_1 & \sum x_8 \\ \sum x_2 & \sum x_2 x_2 & \sum x_3 x_2 & \sum x_4 x_2 & \sum x_5 x_2 & \sum x_6 x_2 & \sum x_1 x_2 & \sum x_8 x_2 \\ \sum x_3 & \sum x_2 x_3 & \sum x_3 x_3 & \sum x_4 x_3 & \sum x_5 x_3 & \sum x_6 x_3 & \sum x_1 x_3 & \sum x_8 x_3 \\ \sum x_4 & \sum x_2 x_4 & \sum x_3 x_4 & \sum x_4 x_4 & \sum x_5 x_4 & \sum x_6 x_4 & \sum x_1 x_4 & \sum x_8 x_4 \\ \sum x_5 & \sum x_2 x_5 & \sum x_3 x_5 & \sum x_4 x_5 & \sum x_5 x_5 & \sum x_6 x_5 & \sum x_1 x_5 & \sum x_8 x_5 \\ \sum x_6 & \sum x_2 x_6 & \sum x_3 x_6 & \sum x_4 x_6 & \sum x_5 x_6 & \sum x_6 x_6 & \sum x_1 x_6 & \sum x_8 x_6 \\ \sum x_7 & \sum x_2 x_7 & \sum x_3 x_7 & \sum x_4 x_7 & \sum x_5 x_7 & \sum x_6 x_7 & \sum x_1 x_7 & \sum x_8 x_7 \\ \sum x_8 & \sum x_2 x_8 & \sum x_3 x_8 & \sum x_4 x_8 & \sum x_5 x_8 & \sum x_6 x_8 & \sum x_1 x_8 & \sum x_8 x_8 \end{vmatrix}}{\Delta} \quad (\text{A.1.26})$$

$$a_8 = \frac{\begin{vmatrix} \theta & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 & \sum x_6 & \sum x_7 & \sum x_1 \\ \sum x_2 & \sum x_2 x_2 & \sum x_3 x_2 & \sum x_4 x_2 & \sum x_5 x_2 & \sum x_6 x_2 & \sum x_7 x_2 & \sum x_1 x_2 \\ \sum x_3 & \sum x_2 x_3 & \sum x_3 x_3 & \sum x_4 x_3 & \sum x_5 x_3 & \sum x_6 x_3 & \sum x_7 x_3 & \sum x_1 x_3 \\ \sum x_4 & \sum x_2 x_4 & \sum x_3 x_4 & \sum x_4 x_4 & \sum x_5 x_4 & \sum x_6 x_4 & \sum x_7 x_4 & \sum x_1 x_4 \\ \sum x_5 & \sum x_2 x_5 & \sum x_3 x_5 & \sum x_4 x_5 & \sum x_5 x_5 & \sum x_6 x_5 & \sum x_7 x_5 & \sum x_1 x_5 \\ \sum x_6 & \sum x_2 x_6 & \sum x_3 x_6 & \sum x_4 x_6 & \sum x_5 x_6 & \sum x_6 x_6 & \sum x_7 x_6 & \sum x_1 x_6 \\ \sum x_7 & \sum x_2 x_7 & \sum x_3 x_7 & \sum x_4 x_7 & \sum x_5 x_7 & \sum x_6 x_7 & \sum x_7 x_7 & \sum x_1 x_7 \\ \sum x_8 & \sum x_2 x_8 & \sum x_3 x_8 & \sum x_4 x_8 & \sum x_5 x_8 & \sum x_6 x_8 & \sum x_7 x_8 & \sum x_1 x_8 \end{vmatrix}}{\Delta} \quad (\text{A.1.27})$$

Hence, the correlation coefficients for the data in Table A.1 are,

$$a_1 = 3,76$$

$$a_2 = 1,75 \times 10^{-4}$$

$$a_3 = 2,00 \times 10^{-4}$$

$$a_4 = 4,28 \times 10^{-5}$$

$$a_5 = 0,42$$

$$a_6 = 0,18$$

$$a_7 = 0,41$$

$$a_8 = 0,16$$

APPENDIX B

The program code written in R programming language with e1071 library for applying multiple regression analysis and SVR to any given data is shown in Syntax 2.

Syntax 2

```
#Calling libraries
library(xlsx)
library(e1071)

#Setting working directory
setwd("C:/datasets")

#Importing data from Excel file
dataset <- read.xlsx("data.xls",1)

#Viewing statistics of the loaded dataset
summary(dataset)
var(dataset$x1)
var(dataset$x2)
var(dataset$x3)
var(dataset$x4)
var(dataset$x5)
var(dataset$x6)
var(dataset$x7)
var(dataset$x8)

#Partitioning dataset into train and test subgroups
dataset.train <- dataset[dataset$subset=="train",2:ncol(dataset)]
dataset.test <- dataset[dataset$subset=="test",2:ncol(dataset)]
print(dataset.train)
print(dataset.test)

#Calculating RSSdefault
def.pred <- mean(dataset.train$x1)
def.rss <- sum((dataset.test$x1-def.pred)^2)
print(def.rss)

#####
#Performing multiple regression analysis#
#####

reg <- lm(x1 ~., data = dataset.train)
print(summary(reg))
reg.pred <- predict(reg,newdata = dataset.test)
reg.rss <- sum((dataset.test$x1-reg.pred)^2) #Calculating RSSmodel
print(reg.rss)
print(1.0-reg.rss/def.rss) #Calculating Pseudo-R2
print(reg.pred)
print(exp(reg.pred))

#####
#Performing Epsilon-SVR w/ Linear kernel#
#####

#10-fold Cross Validation

obj <- tune.svm(x1 ~.,data = dataset.train, scale = T, type = "eps-regression",
               kernel = "linear", cost = seq(from=0.005,to=10.0,by=0.005),epsilon=0.1,
               tolerance=0.001, shrinking=T, fitted=T)
print(obj)
k <- which.min(obj$performances[,3])
c <- obj$performances[,1][k]
RMSE <- obj$performances[,3]
Cost <- obj$performances[,1]
plot(Cost,RMSE,type="l")
```

```

#Regression
epsilon.svr <- svm(x1 ~.,data = dataset.train, scale = T, type = "eps-
regression", kernel = "linear", cost = c,
epsilon=0.1,tolerance=0.001,shrinking=T,fitted=T)
print(epsilon.svr)
esvr2.pred <- predict(epsilon.svr,newdata = dataset.test)
esvr2.rss <- sum((dataset.test$x1-esvr2.pred)^2) #Calculating RSSmodel
print(esvr2.rss)
print(1.0-esvr2.rss/def.rss) #Calculating Pseudo-R2
print(esvr2.pred)
print(exp(esvr2.pred))

#####
#Performing Epsilon-SVR w/ RBF kernel#
#####

#10-fold Cross Validation
obj <- tune.svm(x1 ~.,data = dataset.train, scale = T, type = "eps-regression",
kernel = "radial", cost = seq(from=0.005,to=10.0,by=0.005),epsilon=0.1,
tolerance=0.001, shrinking=T, fitted=T)
print(obj)
k <- which.min(obj$performances[,3])
c <- obj$performances[,1][k]
RMSE <- obj$performances[,3]
Cost <- obj$performances[,1]
plot(Cost,RMSE,type="l")

#Regression
epsilon.svr <- svm(x1 ~.,data = dataset.train, scale = T, type = "eps-
regression", kernel = "radial", cost = c,
epsilon=0.1,tolerance=0.001,shrinking=T,fitted=T)
print(epsilon.svr)
esvr3.pred <- predict(epsilon.svr,newdata = dataset.test)
esvr3.rss <- sum((dataset.test$x1-esvr3.pred)^2) #Calculating RSSmodel
print(esvr3.rss)
print(1.0-esvr3.rss/def.rss)#Calculating Pseudo-R2
print(esvr3.pred)
print(exp(esvr3.pred))

#####
#Performing Nu-SVR w/ Linear kernel#
#####

#10-fold Cross Validation
obj <- tune.svm(x1 ~.,data = dataset.train, scale = T, type = "nu-regression",
kernel = "linear", cost = seq(from=0.005,to=10.0,by=0.005),nu=0.5,
tolerance=0.001, shrinking=T, fitted=T)
print(obj)
k <- which.min(obj$performances[,3])
c <- obj$performances[,1][k]
RMSE <- obj$performances[,3]
Cost <- obj$performances[,1]
plot(Cost,RMSE,type="l")

#Regression
nu.svr <- svm(x1 ~.,data = dataset.train, scale = T, type = "nu-regression",
kernel = "linear", cost = c,
nu=0.5,tolerance=0.001,shrinking=T,fitted=T)
print(nu.svr)
nusvr2.pred <- predict(nu.svr,newdata = dataset.test)
nusvr2.rss <- sum((dataset.test$x1-nusvr2.pred)^2) #Calculating RSSmodel
print(nusvr2.rss)
print(1.0-nusvr2.rss/def.rss) #Calculating Pseudo-R2
print(nusvr2.pred)
print(exp(nusvr2.pred))

#####
#Performing Nu-SVR w/ RBF kernel#
#####

#10-fold Cross Validation
obj <- tune.svm(x1 ~.,data = dataset.train, scale = T, type = "nu-regression",

```

```

        kernel = "radial", cost = seq(from=0.005,to=10.0,by=0.005),nu=0.5,
        tolerance=0.001, shrinking=T, fitted=T)
print(obj)
k <- which.min(obj$performances[,3])
c <- obj$performances[,1][k]
RMSE <- obj$performances[,3]
Cost <- obj$performances[,1]
plot(Cost,RMSE,type="l")

#Regression
nu.svr <- svm(x1 ~.,data = dataset.train, scale = T, type = "nu-regression",
             kernel = "radial", cost = c,
             nu=0.5,tolerance=0.001,shrinking=T,fitted=T)
print(nu.svr)
nusvr3.pred <- predict(nu.svr,newdata = dataset.test)
nusvr3.rss <- sum((dataset.test$x1-nusvr3.pred)^2) #Calculating RSSmodel
print(nusvr3.rss)
print(1.0-nusvr3.rss/def.rss) #Calculating Pseudo-R2
print(nusvr3.pred)
print(exp(nusvr3.pred))

```

CURRICULUM VITAE

Name Surname: Korhan KOR
Place and Date of Birth: İzmir, 25.07.1988
E-Mail: korhank88@gmail.com



EDUCATION:

M.Sc.: Istanbul Technical University (2012-2015)
Department of Petroleum and Natural Gas Engineering
B.Sc.: Istanbul Technical University (2006-2012)
Department of Petroleum and Natural Gas Engineering

PROFESSIONAL EXPERIENCE:

2015 - Turkish Petroleum
Drilling Engineer
2013 – 2015 Istanbul Technical University
Research Assistant

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Kor, K.,** Altun, G., 2015: Destek Vektör Makinesi Yöntemi ile Sondaj İlerleme Hızı Optimizasyonu. *IPETGAS 2015 – 20th International Petroleum and Natural Gas Congress and Exhibition of Turkey*, May 27-29, 2015, Ankara, Turkey.